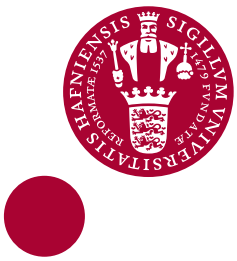


Understanding single-pass transmembrane receptor signaling from a structural viewpoint—what are we missing?

Bugge, Katrine Østergaard; Lindorff-Larsen, Kresten; Kragelund, Birthe Brandt

Published in:
The FEBS Journal

DOI:
10.1111/febs.13793

Publication date:
2016

Document version
Publisher's PDF, also known as Version of record

The FEBS Journal

FEBS PRESS science publishing by scientists

STATE-OF-THE-ART REVIEW

Understanding single-pass transmembrane receptor signaling from a structural viewpoint—what are we missing?

Katrine Bugge, Kresten Lindorff-Larsen and Birthe B. Kragelund

Department of Biology, Structural Biology and NMR Laboratory, University of Copenhagen, Denmark

Keywords
dimerization; NMR; protein structure; signal transduction; single-pass TMDs

Correspondence
B. B. Kragelund, Department of Biology, Structural Biology and NMR Laboratory, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark
Fax: +45 3532 2075
Tel: +45 3532 2081
E-mail: bbk@bio.ku.dk

(Received 19 April 2016, revised 10 June 2016, accepted 27 June 2016)

doi:10.1111/febs.13793

Single-pass transmembrane receptors are involved in essential processes of both physiological and pathological nature and represent more than 1300 proteins in the human genome. Despite the high biological relevance of these receptors, the mechanisms of the signal transductions they facilitate are incompletely understood. One major obstacle is the lack of structures of the transmembrane domains that connect the extracellular ligand-binding domains to the intracellular signaling platforms. Over a period of almost 20 years since the first structure was reported, only 21 of these receptors have become represented by a transmembrane domain structure. This scarceness stands in strong contrast to the significance of these transmembrane α-helices for receptor functionality. In this review, we explore the properties and qualities of the current set of structures, as well as the methodological difficulties associated with their characterization and the challenges left to be overcome. Without an increased and focused effort to bring this class of proteins *on par* with the remaining membrane protein field, a serious lag in their biological understanding looms. Design of pharmaceutical agents, prediction of mutational affects in relation to disease, and deciphering of functional mechanisms require high-resolution structural information, especially when dealing with a domain carrying so much functionality in so few residues.

Introduction

As transmitters of vital molecules and environmental cues to the inside of the membrane-enveloped cell, membrane proteins play indispensable roles for all known life forms. They constitute up to one-third of the proteome [1,2] and are targets for more than 50% of pharmaceutical drugs [3]—still they represent <3% of the structures deposited in the RCSB Protein Data Bank (PDB) [4]. One important type of membrane protein is the single-pass transmembrane receptors (SPTMRs), of which ~ 1300 have been identified in

Abbreviations
APLP, amyloid precursor-like protein; APP, amyloid precursor protein; CD4, cluster determinant 4; CMC, critical micelle concentration; DAP12, lymphoid/myeloid receptor signaling module; DP, detergent per protein; ECD, extracellular domain; Eph, ephrin receptor; EPOR, erythropoitin receptor; ErbB, epidermal growth factor receptor; FGFR, fibroblast growth factor receptor; GHR, growth hormone receptor; GpA, glycophorin A; GP, glycophorin; ICD, intracellular domain; Int, integrin; IR, insulin receptor; JM, juxtamembrane; LP, lipid per protein; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; PDGFR, platelet-derived growth factor receptor; PREs, paramagnetic relaxation enhancements; PRL, prolactin; PRLR, prolactin receptor; RDCs, residual dipolar couplings; RMSD, root-mean-square deviation; RTKs, receptor tyrosine kinases; SPTMRs, single-pass transmembrane receptors; TLR3, Toll-like receptor 3; TMD, transmembrane domain; VEGFR, vascular endothelium growth factor receptor.

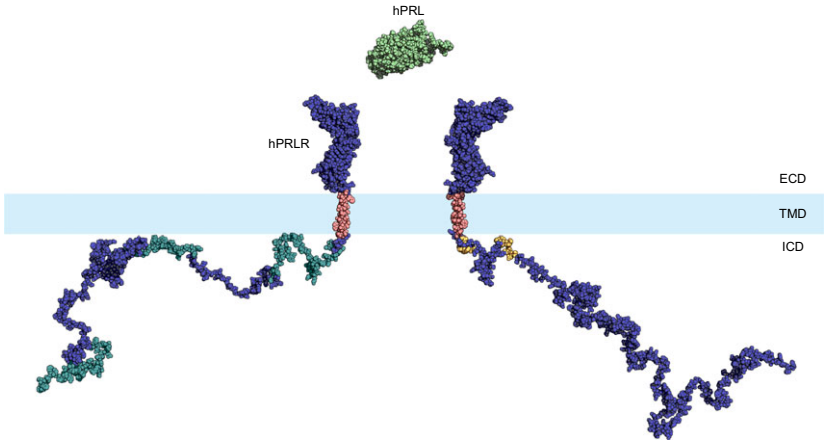


Fig. 1. Topology of an SPTMR, here exemplified by the human PRLR [20]. The membrane-embedded part of the receptor is colored pink, while the water-soluble domains are blue. Two conserved sequence motifs, Box1 and Box2, are orange, and three recently identified lipid interaction domains [14] are cyan. The hormone ligand prolactin (PDB code 1RW5 [108]) is shown in green and marked PRL, while the membrane bilayer is represented by a pale blue box. The active complex is a 1 : 2 complex and the two SPTMRs comprise three main domains: the extracellular domain (ECD), the transmembrane domain (TMD), and the intracellular domain (ICD).

the human genome, constituting roughly 6% of human genes [5]. This membrane protein family includes, e.g., the receptor tyrosine kinases (RTKs), the integrins, and the cytokine receptors, and is involved in a variety of biological processes ranging from metabolism, growth, proliferation, and apoptosis to immune responses [6–9].

The functional signaling unit of SPTMRs is usually assembled from two or more identical or nonidentical receptor chains. These SPTMR chains are composed of several functional regions, divided by the membrane into three distinct domains; the extracellular domain (ECD), the transmembrane domain (TMD), and the intracellular domain (ICD) (Fig. 1). Generally, the ECDs of SPTMRs represent the longest region (up to 22 100 residues with a median of 380 residues), while the ICDs are smaller and less conserved (up to 687 residues with a median of 56 residues) [5]. The TMDs constitute the shortest (~ 25 residues [10]) and most well-conserved domain [11], which connect the ECD and ICD through its unidirectional insertion into the membrane bilayer. The ECDs may be considered the transceivers of the SPTMRs and have the unique ability to recognize large first messengers such as hormones, cytokines, extracellular matrix proteins, and other SPTMRs [5]. The ECDs are globular, soluble domains and are the best characterized, while recent studies have highlighted the high proportion of structural disorder in the ICDs of human SPTMRs [12–14]. As a response to extracellular ligand interactions, these ICDs are modified, leading to initiation of cellular signaling cascades. The majority of SPTMRs does not possess intrinsic enzymatic activity, and instead rely on the recruitment of adaptors and enzymes to propagate signals [5].

The details of cross-membrane signal transduction through SPTMRs remain largely unknown. Nonetheless, SPTMRs have been suggested to function by adopting multiple conformations through ligand-induced stabilization of specific homo- or heterodimeric conformations [15–18]. The high conformational flexibility linked to this mode of action is a fascinating trait of these receptors, allowing delicate local rearrangements to govern the behavior of large and complex signaling pathways. However, this high conformational flexibility along with their diverse structural composition also makes SPTMRs notoriously difficult to characterize structurally [19], and no high-resolution structure of any intact SPTMR has been solved to date. As an alternative, a divide-and-conquer approach is often evoked in structural studies of these proteins, which may be combined with computational modeling [17,20]. Even though a divide-and-conquer strategy should always be applied with care, it is justified for the TMDs by the two-stage (three-step) model of membrane protein folding [21], highlighting that TMD α-helices are independently stable domains, as well as several studies demonstrating that TMDs may preserve their native fold when separated from adjacent domains [22–24].

Accumulating evidence suggests that the membrane-embedded TMDs of SPTMRs regulate receptor chain associations as well as cross-membrane signal transduction by changing conformation or oligomerization

status [15,18,19,25,26], placing the single-pass TMDs in the center of cross-membrane SPTMR functionality. This essential role is further highlighted by the discovery of mutations in the TMDs or the TMD-juxtamembrane (JM) boundary regions that are associated with severe diseases [7,27,28]. Furthermore, several studies have demonstrated how the activity of specific SPTMRs may be controlled by small peptides that specifically recognize their TMDs, and thereby interfere with their lateral association within the membrane [29–31]. Due to this decisive role in receptor functionality and the conserved nature of the TMDs, SPTMR-TMDs represent a promising class of pharmacological targets. Perplexingly, the number of atomic resolution structures of SPTMR-TMDs has remained persistently low, and the functionality and molecular details of these domains therefore remain enigmatic. Currently, this leaves the *modus operandi* of SPTMRs unresolved.

In an attempt to identify the missing pieces in the puzzle of transmembrane signaling, we here assess the structural landscape of mammalian SPTMR-TMDs and discuss the progress and challenges revealed by the limited number of structural studies conducted over the last two decades. We mainly focus on the monomeric and homodimeric TMD structures, the transition between them and their interactions, which have implications for receptor assembly as well as functionality. An increasing effort in the field within the last 5 years has brought the total number of unique SPTMR-TMDs represented by monomeric and/or homodimeric structures in the PDB up to 21 (represented by 31 PDB structures). Structures of three additional SPTMR-TMDs have been deposited in the PDB (2M3E, 2MIC, 2MJO), but as no information regarding the details of, e.g., data acquisition and structure calculation methods have been published, they are not considered further in this review. This means that <2% of the TMDs from this class of membrane proteins have been structurally characterized, and the diversity of the receptor families represented is consequently limited. The majority of the solved structures belongs to the RTKs [the ephrin receptor (Eph), the epidermal growth factor receptor (ErbB), the fibroblast growth factor receptor (FGFR), the vascular endothelium growth factor receptor (VEGFR), and the platelet-derived growth factor receptor (PDGFR) families], while a few immunoreceptors, integrins, and class I cytokine receptors are represented along with one member of the glycophorin (GP) and amyloid precursor-like protein (APLP) families, respectively. This review will highlight that the road toward a full, structure-based mechanistic understanding of SPTMR-TMDs is still long and littered with obstacles.

Single-pass TMDs and their structures

Currently, the dominating methods for atomic resolution structural studies of proteins are X-ray crystallography, solution-state nuclear magnetic resonance (NMR) spectroscopy, solid-state NMR spectroscopy and electron microscopy. While ~ 90% of all structures deposited in the PDB have been solved by X-ray crystallography [32], solution-state NMR spectroscopy has exclusively been used to determine the 31 published structures of monomeric and homodimeric SPTMR-TMDs. The first SPTMR-TMD structure was solved in 1997 of the glycophorin A (GpA) TMD in its homodimeric form [33], demonstrating the applicability of solution-state NMR spectroscopy to this particular domain of SPTMRs. The reason for the dominance of solution-state NMR spectroscopy is probably its high suitability for characterizing small, dynamic proteins, in combination with the challenges associated with crystallizing membrane proteins. Nonetheless, solution-state NMR spectroscopy imposes very specific requirements. First, the protein of interest should be isotope-labeled, currently essentially restricting production to entail recombinant expression in *E. coli* or potentially in yeast or by cell-free synthesis [34]. Furthermore, solution-state NMR spectroscopy is restricted to small-sized systems (below ~ 100 kDa), limiting the selection of membrane mimetics significantly. Within the currently feasible size range is, e.g., organic solvents, detergents, and bicelles, where the latter two represent the most widely used systems to solubilize membrane proteins for NMR studies. Several excellent reviews are available describing the broader palette of membrane-mimicking solvents for solution-state NMR spectroscopy [35–39] as well as detailing the process of structure determination of membrane proteins by this technique [40–43]. Importantly, it is well established that membrane mimetics may influence the quality of the collected data as well as the structural properties of membrane proteins significantly [44], and it is likely that certain types of helix–helix interactions are exaggerated or reduced in membrane mimetics of varying types [45]. Currently, no robust method for predicting which membrane-mimicking solvent will be most suitable for a specific membrane protein exist; this parameter needs to be optimized in each individual case. Evaluating the quality of the NMR data of the protein in various membrane mimetics is, however, relatively straightforward from analysis of ^{1}H,^{15}N-HSQC spectra. When it comes to evaluating whether the structure is perturbed under the applied conditions relative to the native structure, the most reliable measure is conservation of biological activity. However, in case of SPTMR-

TMDs, biological activity is currently only measurable in the context of the full receptor within the cell. For example, ligand binding to the ECD may be uncoupled from signaling to the ICD, and *vice versa*, demonstrating that such biochemical assays do not guarantee the TMD to exist in the correct signaling-competent conformation. Thus, this approach is critically unfeasible as a screening methodology for evaluating the SPTMR-TMDs in different membrane mimetics compatible with NMR studies. Adding to this, SPTMR-TMDs have a higher surface-to-volume ratio than multi-pass membrane proteins, likely causing an increased sensitivity to the surrounding solvent. Use of membrane mimetics in structural studies of these domains therefore requires careful evaluation of putative effects on the structure and obtained data.

Monomeric TMD structures

The TMD of 10 SPTMRs is represented by a monomeric structure in the PDB (Fig. 2). Of these, three belong to the integrin (Int) family (IntαI [46], Intβ3 [47], IntαIIβ [48]), three are RTKs (insulin receptor (IR) [49], ErbB1[50], ErbB2 [51]), two are class I cytokine receptors (the erythropoitin receptor (EPOR) [52,53] and the prolactin receptor (PRLR) [20]), one member is an immunoreceptor (cluster determinant 4 (CD4) [54]), and one belongs to the APLP family (amyloid precursor protein (APP) [55,56]). The EPOR is represented by two similar murine (2MXB [53]) and human (2MV6 [52]) TMD structures, while the APP-TMD structure has been determined by two different groups (2LLM [55] and 2LP1 [56]). All the monomeric structures consist of one main α-helix harboring 24–35 residues, encompassing the region bioinformatically predicted to be the TMD [57]. The TMD structures of CD4, ErbB1, and mErbB2 also have minor C-terminal JM α-helices of 13, 9, and 7 residues, respectively, connected to the main helix by 10 (CD4)- or 3-residue loop regions (Fig. 2). The mEPOR-TMD and the two APP-TMD structures have minor N-terminal α-helices that are between 7 and 10 residues, connected to the main helix by a short loop region (Fig. 2).

The majority of the monomeric TMD structures appear to have some degree of curvature, with only the EPOR-TMDs, the Intβ3-TMD, and the IntαIIβ-TMD represented by completely straight α-helices (Fig. 2). Excluding one of the APP-TMD structures (2LP1), the remainder has various degrees of subtle kinks only slightly changing the direction of the helix axis. However, one of the APP-TMD structures (2LP1) has a significant kink near the center of the TMD helix of ~ 30°, mainly caused by a Gly-pair. The curved nature of this TMD is believed to be crucial for recognition and proteolysis by γ-secretase [56]. The other structure of the APP-TMD monomer (2LLM) only possesses a slight bend at the same position.

Homodimeric TMD structures

The TMDs of 14 SPTMRs are represented by a homodimer structure in the PDB (Fig. 3). Nine of

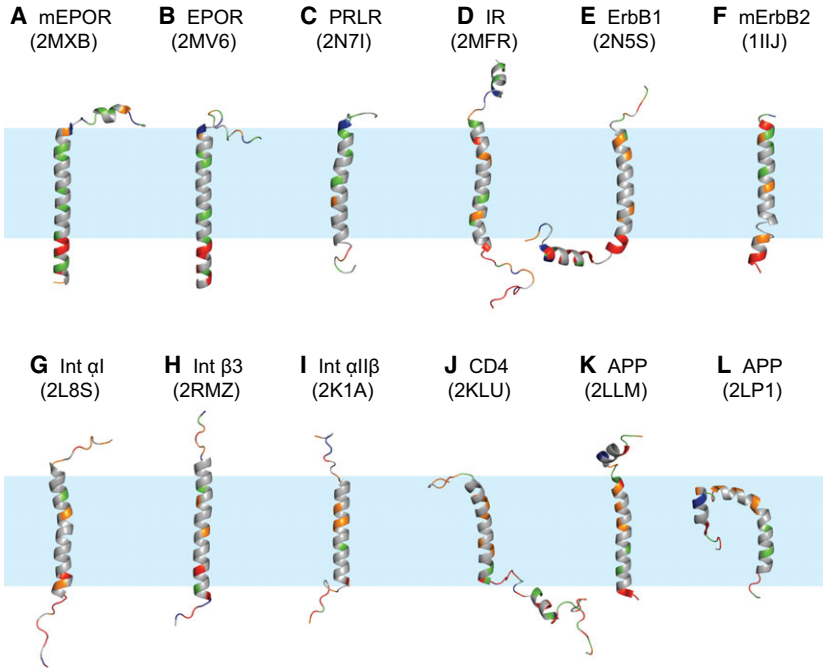


Fig. 2. Currently available monomeric structures of mammalian SPTMR-TMDs. The pale blue box represents the membrane bilayer hydrocarbon layer of ~ 30 Å. Residues with hydrophobic side chains are gray, polar are green, positively charged are red, negatively charged are blue. Cys, Gly, and Pro are orange. (A) mouse (m) EPOR-TMD (2MXB), (B) EPOR-TMD (2MV6), (C) PRLR-TMD (2N7I), (D) IR-TMD (2MFR), (E) ErbB1-TMD (2N5S), (F) mErbB2-TMD (1IIJ), (G) Integrin αI-TMD (2L8S), (H) Integrin β3-TMD (2RMZ), (I) Integrin αIIβ-TMD (2K1A), (J) hCD4-TMD (2KLU), (K) APP-TMD (2LLM), (L) APP-TMD (2LP1).

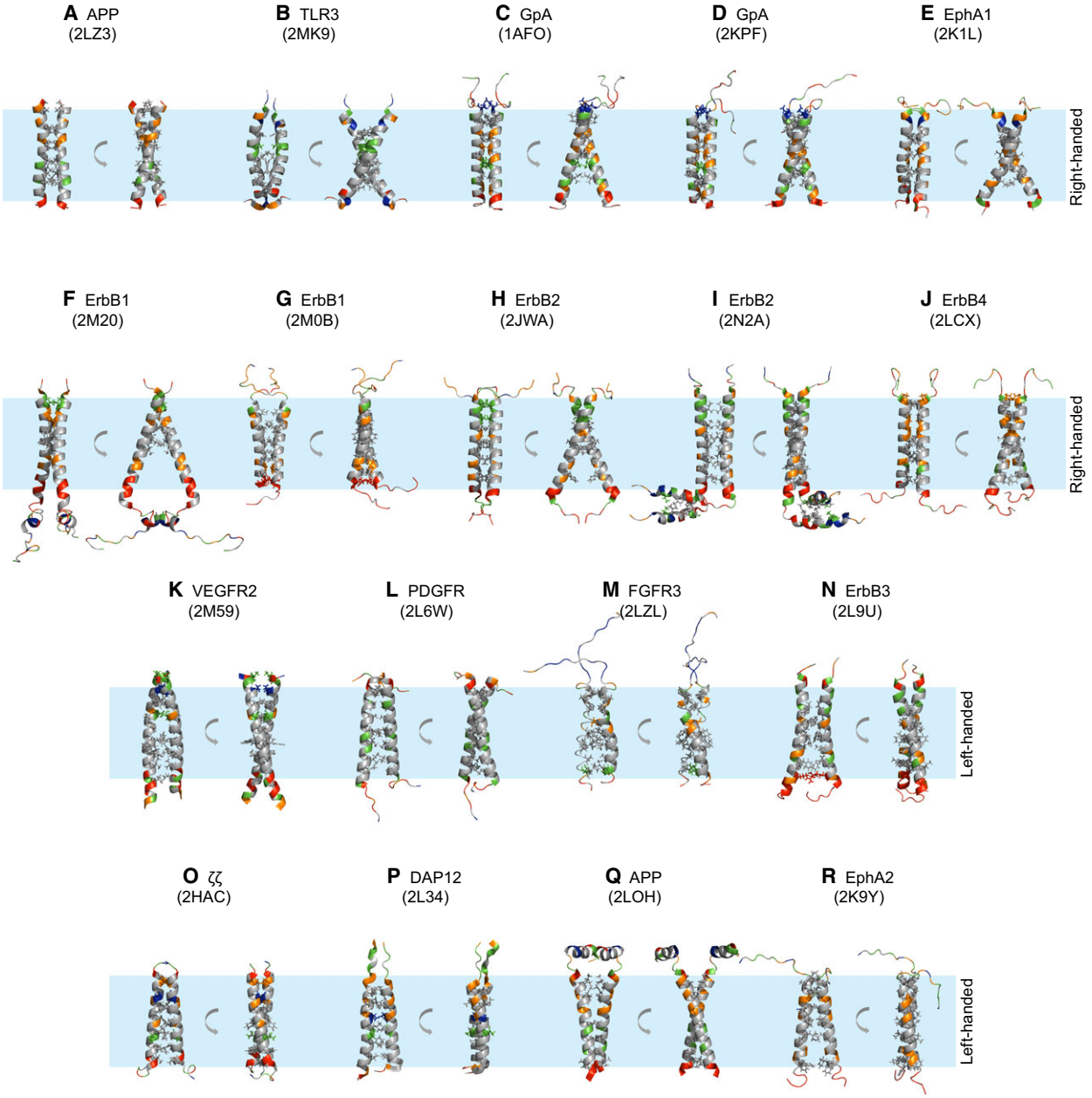


Fig. 3. Currently available homodimer structures of mammalian WT SPTMR-TMDs. The right-handed structures are shown in the two top panels, while the left-handed structures are shown in the two bottom panels. The pale blue box represents the membrane bilayer hydrocarbon layer of ~ 30 Å. Residues with hydrophobic side chains are gray, polar are green, positively charged are red, negatively charged are blue. Cys, Gly, and Pro are orange. Residues involved in interhelical interactions are shown as sticks. (A) APP-TMD (2LZ3), (B) TLR3-TMD (2MK9), (C) GpA-TMD (1AFO), (D) GpA-TMD (2KPF), (E) EphA1-TMD (2K1L), (F) ErbB1-TMD (2M20), (G) ErbB1-TMD (2M0B), (H) ErbB2-TMD (2JWA), (I) ErbB2-TMD (2N2A), (J) ErbB4-TMD (2LCX), (K) VEGFR2-TMD (2M59), (L) PDGFR-TMD (2L6W), (M) FGFR3-TMD (2LZL), (N) ErbB3-TMD (2L9U), (O) ζζ-TMD (2HAC), (P) DAP12-TMD (2L34), (Q) APP-TMD (2LOH), (R) EphA2 (2K9Y).

these are RTKs (ErbB1-4 [15,58–62], FGFR3 [63], PDGFR [64], VEGFR2 [23], EphA1-2 [65,66]), three are immunoreceptors (lymphoid/myeloid receptor signaling module (DAP12) [67], the ζ chain (ζζ) [68], Toll-like receptor 3 (TLR3) [69]), one belongs to the APLPs (APP [70,71]), and one belongs to the GP family (GpA [33], [72]). Two similar homodimer structures exist of the GpA-TMD and the APP-TMD (1AFO [33] and 2KPF/2KPE [72] for the GpA and 2LOH [71] and 2LZ3 [70] for the APP), which in each case have

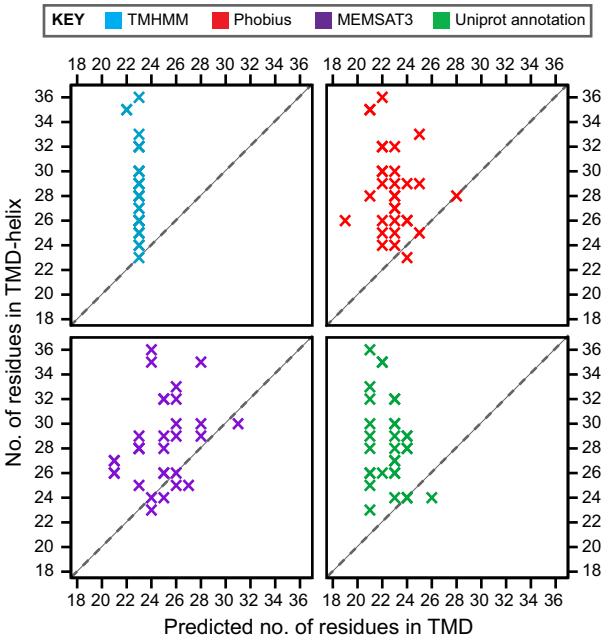


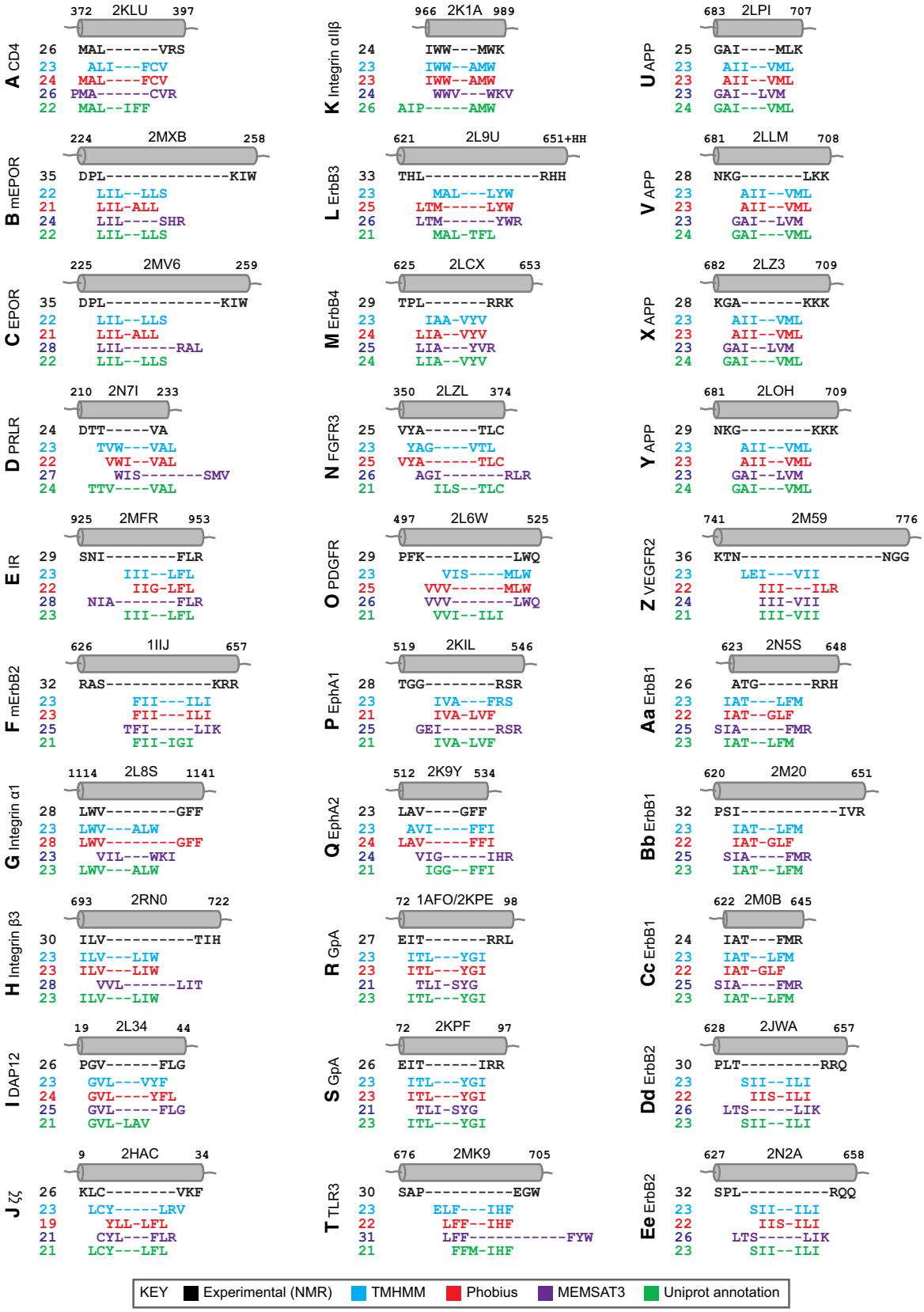
Fig. 4. Number of residues in the TMD α-helix of the NMR structures plotted against the bioinformatically predicted number of residues in the TMD. The results from four different predictions have been used as x-coordinates; TMHMM [73] (light blue), Phobius [75,76] (red), MEMSAT3 [77] (purple), and Uniprot annotations (green) [57,74], while the number of residues in the TMD α-helix of the NMR structure is the y-coordinate. NMR structures represented by a cross located above the dashed line have longer α-helices than predicted, and *vice versa*.

been solved by different groups, while two different homodimer structures have been solved of the ErbB1-TMD (2M20 [15] and 2M0B [62]) and the ErbB2-TMD (2JWA [58] and 2N2A [61]).

The TMD homodimer structures all consist of two main α-helices harboring 23–38 residues each, and include the regions bioinformatically predicted to be the TMD [57]. The ErbB1-TMD structure 2M20 and the ErbB2-TMD structure 2N2A have additional minor C-terminal JM α-helices of 11 and 15 residues, respectively, connected to the main helix by a three-residue loop (Fig. 3). Like the monomeric structures, the majority of the homodimeric structures have subtle kinks or concave helices, while four structures appear to have straight helices (EphA1-TMD, EphA2-TMD, ErbB2-TMD, GpA-TMD) (Fig. 3). The VEGFR2-TMD structure represents a special case having two convex helices, resulting in a slight bending of the two helices moving through their entire length. The FGFR3-TMD structure stands out by forming a concave helix in the bioinformatically predicted TMD region, interrupted by an N-terminal kink before another four residues resume the helix. Likewise, the APP-TMD structure (2LZ3) has both an N-terminal and C-terminal kink. The remaining structures have one minor kink each, but none as severe as the monomeric APP-TMD structure (2LP1). The majority of these kinks occur in proximity to Gly residues, while a few are adjacent to Ser and/or Thr.

Extent of the TMD α-helix

Upon closer inspection of the monomeric and homodimeric SPTMR-TMD structures, we noted that the Uniprot annotated TMD regions [57] often only partially describe the region that was experimentally determined to be helical. Some of the more pronounced examples are the mEPOR-TMD and the hEPOR-TMD in which 10 additional helical residues extent the annotated TMD region C-terminally, while the ErbB1-TMD structure 2M20, the ErbB3-TMD structure, and the VEGFR2-TMD structure are extended by a total of 9, 12, and 16 residues, respectively. In the full protein, some of the TMDs may have even longer α-helices than experimentally determined, as the TMD constructs have been terminated close to the observed helix-border (e.g., the VEGFR2-TMD which is α-helical throughout its length). The Uniprot annotation of TMD regions is performed by applying the TM helix predictor TMHMM [73] as a Yes/No criterion. The sequence positions are subsequently assigned by the hydrophobic moment plot method of Eisenberg *et al.* [74] providing an average TMD length of 21 residues. However, the Uniprot annotated TMD regions are on average ~ 20% shorter than the TMD α-helix of the NMR structures. To further understand this observation, the complete receptor sequences including the extra- and intracellular parts were submitted directly to the TMHMM- [73], Phobius- [75,76], and MEMSAT3 [77] webservers (Figs 4 and 5), which are three commonly applied secondary structure and topology predictors for all-helical transmembrane proteins. Overall, the results of these predictors suggested shorter α-helices than was observed in the NMR structures (Fig. 4), with the MEMSAT3 predictor seemingly providing the results overall closets to the NMR structures. When the results of the predictors were inspected more closely, it became evident that the predicted TMD α-helices were often shorter C-terminally compared to the NMR structures (Fig. 5). It should additionally be noted that TMHMM suggested a second TMD region for the mErbB2, ErbB1, ErbB2, ErbB4, FGFR3, and VEGFR2, which if correct would place their intracellular domains on the extracellular side.


A CD4
372 2KLU 397
26 MAL------VRS
23 ALI---FCV
24 MAL----FCV
26 PMA------CVR
22 MAL--IFF
B mEPOR
224 2MXB 258
35 DPL--------------KIW
22 LIL--LLS
21 LIL-ALL
24 LIL----SHR
22 LIL--LLS
C EPOR
225 2MV6 259
35 DPL-------------KIW
22 LIL--LLS
21 LIL-ALL
28 LIL------RAL
22 LIL--LLS
D PRLR
210 2N7I 233
24 DTT-----VA
23 TVW---VAL
22 VWI--VAL
27 WIS-------SMV
24 TTV----VAL
E IR
925 2MFR 953
29 SNI--------FLR
23 III--LFL
22 IIG-LFL
28 NIA-------FLR
23 III--LFL
F mErbB2
626 1IIJ 657
32 RAS-----------KRR
23 FII---ILI
23 FII---ILI
25 TFI-----LIK
21 FII-IGI
G Integrin α1
1114 2L8S 1141
28 LWV--------GFF
23 LWV---ALW
28 LWV--------GFF
23 VIL---WKI
23 LWV---ALW
H Integrin β3
693 2RN0 722
30 ILV----------TIH
23 ILV---LIW
23 ILV---LIW
28 VVL------LIT
23 ILV---LIW
I DAP12
19 2L34 44
26 PGV------FLG
23 GVL---VYF
24 GVL----YFL
25 GVL-----FLG
21 GVL-LAV
J ζζ
9 2HAC 34
26 KLC--------VKF
23 LCY-----LRV
19 YLL-LFL
21 CYL---FLR
21 LCY---LFL
K Integrin αIIβ
966 2K1A 989
24 IWW---MWK
23 IWW--AMW
23 IWW--AMW
24 WWV---WKV
26 AIP-----AMW
L ErbB3
621 2L9U 651+HH
33 THL-----------RHH
23 MAL---LYW
25 LTM-----LYW
26 LTM------YWR
21 MAL-TFL
M ErbB4
625 2LCX 653
29 TPL-------RRK
23 IAA-VYV
24 LIA--VYV
25 LIA---YVR
24 LIA--VYV
N FGFR3
350 2LZL 374
25 VYA------TLC
23 YAG----VTL
25 VYA------TLC
26 AGI-------RLR
21 ILS--TLC
O PDGFR
497 2L6W 525
29 PFK----------LWQ
23 VIS----MLW
25 VVV------MLW
26 VVV-------LWQ
21 VVI--ILI
P EphA1
519 2KIL 546
28 TGG--------RSR
23 IVA---FRS
21 IVA-LVF
25 GEI------RSR
21 IVA-LVF
Q EphA2
512 2K9Y 534
23 LAV----GFF
23 AVI----FFI
24 LAV-----FFI
24 VIG-----IHR
21 IGG--FFI
R GpA
72 1AFO/2KPE 98
27 EIT-------RRL
23 ITL---YGI
23 ITL---YGI
21 TLI-SYG
23 ITL---YGI
S GpA
72 2KPF 97
26 EIT------IRR
23 ITL---YGI
23 ITL---YGI
21 TLI-SYG
23 ITL---YGI
T TLR3
676 2MK9 705
30 SAP----------EGW
23 ELF---IHF
22 LFF--IHF
31 LFF----------FYW
21 FFM-IHF
U APP
683 2LPI 707
25 GAI----MLK
23 AII--VML
23 AII--VML
23 GAI--LVM
24 GAI---VML
V APP
681 2LLM 708
28 NKG-------LKK
23 AII--VML
23 AII--VML
23 GAI--LVM
24 GAI---VML
X APP
682 2LZ3 709
28 KGA-------KKK
23 AII--VML
23 AII--VML
23 GAI--LVM
24 GAI---VML
Y APP
681 2LOH 709
29 NKG--------KKK
23 AII--VML
23 AII--VML
23 GAI--LVM
24 GAI---VML
Z VEGFR2
741 2M59 776
36 KTN----------------NGG
23 LEI---VII
22 III---ILR
24 III-VII
21 III-VII
Aa ErbB1
623 2N5S 648
26 ATG------RRH
23 IAT---LFM
22 IAT--GLF
25 SIA-----FMR
23 IAT--LFM
Bb ErbB1
620 2M20 651
32 PSI----------IVR
23 IAT--LFM
22 IAT-GLF
25 SIA----FMR
23 IAT--LFM
Cc ErbB1
622 2M0B 645
24 IAT---FMR
23 IAT--LFM
22 IAT-GLF
25 SIA----FMR
23 IAT--LFM
Dd ErbB2
628 2JWA 657
30 PLT---------RRQ
23 SII--ILI
22 IIS-ILI
26 LTS-----LIK
23 SII--ILI
Ee ErbB2
627 2N2A 658
32 SPL----------RQQ
23 SII--ILI
22 IIS-ILI
26 LTS-----LIK
23 SII--ILI
KEY Experimental (NMR) TMHMM Phobius MEMSAT3 Uniprot annotation

The origin of the discrepancy between the results of the applied secondary structure and topology predictors and the NMR structures of the SPTMR-TMDs remains to be understood. However, it seems that especially Uniprot and the TMHMM predictor operate with a very narrow definition of the length of a membrane spanning helix (21–23 residues, Fig. 4), which intuitively would be expected to vary with, e.g., cell membrane composition and helix tilt angle. Another potential issue may reside in distortions of the TMD peptides brought on by the membrane mimetics, or simply that the α-helical region is not restricted to the hydrocarbon core, and could extend into the polar regions of the bilayer as well as into the inside and outside of the cell. In this regard, it should be remembered that the ~ 30 Å hydrocarbon core of a membrane bilayer is surrounded by two head group regions each having a thickness of ~ 15 Å with no sharp boundaries [38]. Interestingly, the different structures determined of the ErbB1-TMD (2M20, 2M0B, 2N5S) vary considerably in the extent of the TMD α-helix. The TMD α-helix of the monomeric (2N5S) and dimeric (2M0B) structures solved in DPC micelles harbor 26 and 25 residues, respectively, compared to 32 residues of the dimeric structure solved in bicelles (2M20). This may suggest that the length difference is connected to differences in the membrane mimetics, with the bicelles having a thicker hydrocarbon core which could favor a longer α-helix. On the other hand, the dimeric structures of the ErbB2-TMD (2N2A) and the GpA-TMD (2KPE/1AFO) solved in DPC micelles harbor 31 and 27 residues, respectively, in the TMD α-helix compared to 30 and 26, respectively, in bicelles (2JWA and 2KPF). Of further note, the α-helix of the ErbB3-TMD structure continues into the non-native His-tag placed at its C terminus, suggesting a non-native extension of the helix. Although the origin of the discrepancies in the extent of the TMD α-helix remains to be elucidated, these observations nonetheless have important implications for the design of TMD peptides for, e.g., NMR studies and may suggest problems with the nativeness of the solved SPTMR-TMD structures.

Helix packing of the TMD homodimers

Two α-helices may acquire close packing by interleaving their side chains in the 'knobs into holes' [78] or 'ridges into grooves' [79] arrangements. The steric limitations of these arrangements favor specific crossing angles, and in 1977, Chothia *et al.* [79] introduced the three basic helix packing types defined by crossing angles; (a) a mean angle between the helix axes of − 82°, (b) a mean angle between the helix axes of − 60°, and (c) a mean angle between the helix axes of + 19°. These helix crossing angles are the most favorable for close helix packing, but the observed crossing angles often differ slightly due to variations in side-chain volumes.

The currently solved SPTMR-TMD homodimer structures have an almost equal distribution of type II and III packing angles (Fig. 6, Table 1), while no examples of type I packing angles have been published. The current structures may therefore be divided into two main groups; group II containing the homodimers with type II helix packing angles resulting in right-handed crossing, and group III having type III helix packing angles resulting in left-handed crossing. It is, however, noted that three right-handed structures (2LZ3, 2M0B, 2N2A) have a helix crossing angle which is significantly higher than the remainder of the group (Fig. 6, highlighted with '*').

The ideal group II homodimers are the TMD structures of TLR3, EphA1, ErbB1 (2M20), ErbB2 (2JWA), ErbB4, and GpA (Table 1, Fig. 3B–F,H,J). The dimeric TMD structures of APP (2LZ3), ErbB1 (2M0B), and ErbB2 (2N2A) are denoted II* due to their unusual helix crossing angles (Table 1, Fig. 3A, G,I), and they are therefore considered as a separate subgroup of group II. The ideal group II structures are characterized by a right-handed crossing of their α-helices resulting in a relatively large crossing angle (average crossing angle of − 41°) and smaller contact surfaces (average of ~ 400 Å^2) (Table 1).

Fig. 5. Schematic overview of the extent of the TMD α-helices determined by NMR spectroscopy and predicted bioinformatically. Alignment of the three first and the three last residues of the TMD α-helix as determined by NMR spectroscopy (black), or the predictors TMHMM [73] (light blue), Phobius [75,76] (red), MEMSAT3 [77] (purple), or Uniprot annotations (green) [57,74]. The numbers to the left of the sequences are the number of residues in each sequence. The gray cylinders represent the relative length of the TMD α-helix as determined by NMR spectroscopy with the sequence number of the first and last residue in the TMD α-helix as well as the PDB code given. (A) CD4 (2KLU), (B) mEPOR (2MXB), (C) EPOR (2MV6), (D) PRLR (2N7I), (E) IR (2MFR), (F) mErbB2 (1IIJ), (G) Integrin αI (2L8S), (H) Integrin β3 (2RMZ), (I) DAP12 (2L34), (J) ζζ (2HAC), (K) Integrin αIIβ (2K1A), (L) ErbB3 (2L9U), (M) ErbB4 (2LCX), (N) FGFR3 (2LZL), (O) PDGFR (2L6W), (P) EphA1 (2KIL), (Q) EphA2 (2K9Y), (R) GpA (1AFO/2KPE), (S) GpA (2KPF), (T) TLR3 (2MK9), (U) APP (2LPI), (V) APP (2LLM), (X) APP (2LZ3), (Y) APP (2LOH), (Z) VEGFR2 (2M59), (Aa) ErbB1 (2N5S), (Bb) ErbB1 (2M20), (Cc) ErbB1 (2M0B), (Dd) ErbB2 (2JWA), (Ee) ErbB2 (2N2A).

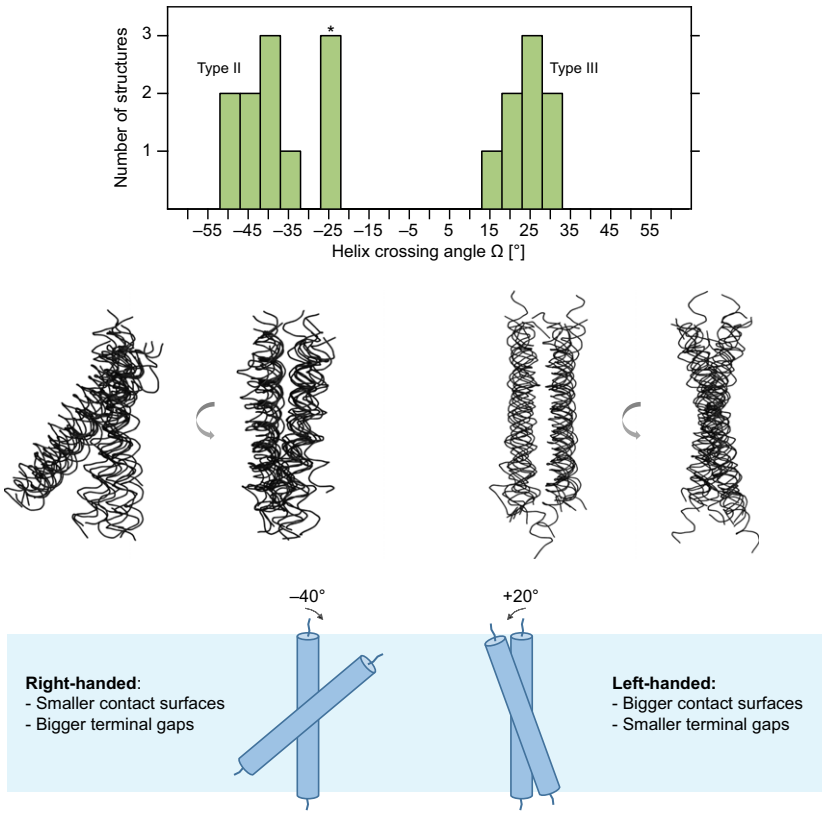


Fig. 6. The two different topology groups of the current SPTMR-TMD homodimer structures. The current SPTMR-TMD homodimer structures can be divided into two groups: group II with type II crossing angles and right-handed helix crossing, and group III with type III crossing angles and left-handed helix crossing. Top: Histogram of the helix crossing angle distribution in the 19 unique homodimer structures. The crossing angles of the APP-TMD homodimer 2LZ3, the ErbB1-TMD homodimer 2M0B, and the ErbB2-TMD homodimer 2N2A are outliers, and their bar is highlighted with '*'. Middle: Backbone alignment of the TMD-α-helices of the left- and right-handed structures (2LZ3, 2M0B and 2N2A left out), respectively, in ribbon representation. Bottom: Schematic representation of the two topology groups.

Consequently, they have a similar overall topology (Fig. 6). Their relatively large crossing angles and small contact surfaces are the result of small-residue motifs commonly encountered in this packing group, which is described further below. In addition, the ideal type II dimers are most separated C-terminally toward the cytoplasmic side, with an average distance between the α-helix C terminus of ~ 20 Å compared to ~ 11 Å for the N terminus. In contrast, the unusual type II* dimers share, except for the right-handedness, many of their characteristics with the left-handed group, such as an almost parallel conformation and bigger contact surfaces.

The group III structures are the TMDs of EphA2, ζζ, APP (2LOH), ErbB3, PDGFR, VEGFR2, FGFR3, and DAP12 (Table 1, Fig. 3K–R). These homodimers are characterized by a left-handed crossing of their α-helices (average crossing angle of 26°), bigger contact surfaces (average of ~ 590 Å^2), and smaller terminal gaps (Table 1), and consequently, like group II, share a similar overall topology (Fig. 6). The α-helical dimers with a left-handed crossing sometimes employ heptad-like dimerization motifs and are also known as 'coiled-coils' [80,81]. It is noted that the two dimer structures of the APP-TMD (2LZ4 and 2LOH) fall into different packing groups, which will be discussed below.

Motifs and interactions in the homodimeric TMD structures

Sequence motifs and the forces driving helix–helix assembly

To ensure specific assembly of the SPTMRs, the TMDs must present specific complementary recognition domains. Although specificity between soluble domains often is achieved by large recognition interfaces (~ 800 Å^2 [82,83]), the small size of single-pass TMDs represents a unique challenge to specificity. The achievement of a sufficient affinity and specificity in these small domains have so far mainly been attributed to recurrent sequence motifs, characterized by specific amino acid patterns. Consequently, an extensive amount of effort has been put into identifying and recognizing motifs in the primary structure of TMDs [27,84–86], and currently, the most studied interaction motifs for TMDs are the so-called GAS and heptad motifs.

The GAS-motif, which is found in ~ 57% of all TMDs [87], is a five-residue small-x$_3$-small residue

Table 1. Characteristics of available WT SPTMR-TMD homodimer structures.

PDB code	Abbrev.	Solvent	Orientation	Motif	Helix crossing angle	Helix packing group	Contact surface area (Å^2)	Dist. N_t (Å)	Dist. C_t (Å)	$N_t - C_t$ (Å)	State	Ref.
2MK9	TLR3	DPC	R	$Fx_3Tx_3LIx_2FIx_2L$	− 51.1	II	380 ± 20	12.7	23.4	− 10.7	OFF	[69]
2K1L	EphA1	Bicelle	R	$AVx_2GLx_2GAx_2LL$	− 50.0	II	530 ± 30	13.2	16.1	− 2.9	ON	[65]
2M20	ErbB1	Bicelle	R	$TGx_2GA\ x_2LLx_2V$	− 44.4	II	N/A	11.6	20.7	− 9.1	ON	[15]
2LCX	ErbB4	Bicelle	R	$PLx_2AGx_2GGx_2ILx_2Vx_3F$	− 43.3	II	360 ± 30	13.1	27.5	− 14.4	ON	[60]
2JWA	ErbB2	Bicelle	R	$Tx_3Sx_2VGx_2LVx_2Lx_3F$	− 41.8	II	360 ± 30	13.8	21.3	− 7.5	ON	[58]
2KPF	GpA	Bicelle	R	$Ex_2LIx_2GVx_2GVx_2TIx_2I$	− 38.0	II	380 ± 20	7.5	20.2	− 12.7	–	[72]
1AFO	GpA	DPC	R	$Ex_2LIx_2GVx_2GVx_2Tx_3I$	− 39.2	II	400	6.9	18.5	− 11.6	–	[33]
2KPE	GpA	DPC	R	$Ex_2LIx_2GVx_2GVx_2TIx_2I$	− 36.9	II	370 ± 20	7.4	21.0	− 13.6	–	[72]
2M0B	ErbB1	DPC	R	$Mx_2ALx_2LLx_2ALx_2GLx_2R$	− 26.8	II*	670 ± 40	14.0	11.9	+ 2.1	OFF	[62]
2LZ3	APP	DPC	R	$Ix_3Mx_2GVx_2Ax_2IVx_2L$	− 25.4	II*	N/A	16.6	13.8	+ 2.8	–	[70]
2N2A	ErbB2	DPC	R	$Ix_3Vx_2LLx_2VLx_2VFx_2L$	− 23.1	II*	420	15.9	21.3	− 5.4	OFF	[61]
2K9Y	EphA2	Bicelle	L	$LAxIGx_2AVxVVx_2LVx_5FF$	+ 15.2	III	490 ± 30	5.6	9.6	− 4.0	–	[66]
2HAC	ζζ	5 : 1 DPC : SDS	L	$Cx_3Dx_2Lx_2Yx_3LTx_2Fx_2V$	+ 20.0	III	576	13.0	16.3	− 3.3	–	[68]
2LOH	APP	DPC	L	$Ix_3Mx_2Gx_3Ax_2Ix_3Lx_2L$	+ 22.6	III	660 ± 60	18.3	10.5	+ 7.8	–	[71]
2L9U	ErbB3	DPC	L	$Ix_2LVxIFx_2Lx_3FLx_2R$	+ 24.9	III	480 ± 40	17.9	18.8	− 0.9	ON	[59]
2L6W	PDGFR	DPC	L	$Vx_3Ax_2Ax_2VLx_2Ix_3Ix_2M$	+ 25.4	III		12.8	14.3	− 1.5	–	[64]
2M59	VEGFR2	DPC	L	$Tx_2Ex_2ILx_2TAx_2Ax_2$ FWxLLx$_2$Ix$_3$V	+ 26.7	III	613 ± 20	11.4	11.8	− 0.4	OFF	[23]
2LZL	FGFR3	9 : 1 DPC : SDS	L	$YAx_2Lx_2Gx_2$ FFx$_2$ILx$_2$Ax$_2$TL	+ 32.1	III	870 ± 40	14.2	15.3	− 1.1	OFF	[63]
2L34	DAP12	LA	L	$Lx_2IVxGDx_2LTx_2Ix_3V$	+ 33.0	III	440 ± 9	7.3	11.9	− 4.6	–	[67]

Abbrev.: Short abbreviated name for the receptor of the TMD. Orientation: R is right-handed and L is left-handed. Motif: residues involved in the interhelix interactions. Helix crossing angle: the angle between the helix directional vectors as defined at the point of crossing. Helix packing group: assignment as defined by the crossing angle. Contact surface area: helix–helix contact surface area as calculated by the authors of the structure. Dist. N_t: average distance between the N terminus of the TMD α-helices in each structural ensemble. Dist. C_t: average distance between the C terminus of the TMD α-helices in each structural ensemble. $N_t - C_t$: difference between the average distances of the N and C termini. A negative difference means that the C-terminus of the TMD α-helices are furthest apart, and *vice versa*. State: signifies if the structure has been assigned to a specific receptor state, here ON for active and OFF for inactive. Ref.: reference describing the structure determination.

motif with the small residue dominantly being G, but occasionally also A or S. The GAS-motif is also known as the GAS_{right}-motif, as it is responsible for the right-handed helical associations of parallel dimers. The Gx_3G version of this motif is the most common [87], and is often found in $i \pm 1$ or $i \pm 2$ association with large aliphatic residues (I, V, and L) [88]. The structural rationale behind the high implication of GAS-motifs in TMD oligomerization is rooted in its propensity for formation of stabilizing interhelical carbon hydrogen bonds, occurring between C^{α}-H^{α} donors and backbone carbonyl oxygen acceptors (C^{α}-$H^{\alpha}\cdots O{=}C$) [89,90]. The small residues in the GAS-motif allow close proximity of the opposing helical backbone, enabling the formation of the interhelical backbone carbon hydrogen bonds, as well as proximity between the main-chain and hydrogen-bonding donors such as Ser [89]. Although Gx_3G is often regarded as a dominating dimerization motif, nonstructural studies on the affinity of TMD dimers harboring Gx_3G motifs indicated that these motifs are neither necessary nor sufficient for TMD dimerization [45,87]. Furthermore, no correlation between the measured dimerization propensity and the presence of GAS-motifs has been found [87], as e.g., established in a study using the TOXCAT assay to investigate the dimerization propensity of 58 RTK TMDs in a bacterial membrane [91].

In the heptad motif, also referred to as the leucine zipper or GAS_{left} motif, the interacting residues form a repeated motif of seven residues (*abcdefg*) [81,92,93]. The heptad motif is responsible for formation of left-handed coiled-coil packing, and in soluble proteins, nonpolar residues are usually found in positions *a* and

d, forming the hydrophobic core of the helix–helix interface, while position *e* and *g* often hold charged residues, forming salt-bridges to each other as well as hydrophobic contacts to the core residues [81,92,93]. However, the name GAS_{left} has been earned as positions *a* and *d* in TMD pairs may also be occupied by small residues (e.g., G, A, S, C) [27]. The *b*, *c*, and *f* positions are on the outside of the interface, in contact with the solvent, which in the case of TMDs is the lipid/detergent. Critics of the TMD-heptad motifs argue that due to the low diversity of amino acids in a TMD region, heptad motifs can almost always be assigned regardless of any true relevance [45]. Furthermore, charged residues, classically found at position *e* and *g*, are rarely found within the membrane-embedded region.

The two motifs described above do, however, not account for all TMD–TMD helix associations. A much more diverse range of physical interactions including van der Waals', aromatic π–π, cation–π, and polar interactions [38,45,94,95] act to stabilize helix–helix associations in the membrane environment, an observation underlined further in the following.

Interaction motifs in the SPTMR-TMD homodimers

Of the 17 homodimer TMD structures in Table 1, 11 are categorized as right-handed (type II/II*) and 8 as left-handed (type III). The residues involved in intermolecular interactions in the dimers are positioned with specific spacings, serving to place them on the same side of the α-helix to form a recognition surface. These appear most frequently to be two-residue spacings (XxxX), less frequently three-residue spacings (XxxxX), one-residue spacings (XxX), and only in a few cases with longer spacing. The average number of interhelical interaction residues in the structures is 8.7 residues per α-helix (7–10 residues) for the right-handed structures, and 9.6 residues per α-helix (7–12 residues) for the left-handed structures.

The right-handed group is dominated by GAS-motifs (employed by 7/9 unique interaction interfaces), but even though these motifs are often highlighted as the defining dimerization motif, other residues are crucial for stabilizing the packing. This role appears primarily to be filled by residues with large, hydrophobic side chains (Leu, Ile, Val, Phe, Met). The right-handed GAS-mediated structures typically have the small-x_3-small motif at the site of helix crossing, ensuring close packing at the crossing site, with a helix crossing angle clustering around − 40° (Fig. 6). The right-handed dimer conformations which do not utilize GAS-motifs are the TLR3-TMD (2MK9) and the ErbB2-TMD (2N2A) structures. In the TLR3-TMD dimer, the residues $Fx_3Tx_3LIx_2FIx_2L$ constitute the helix–helix interface, which is stabilized by van der Waals' interactions between the bulky side chains and stacking of the phenyls [69]. Interestingly, the polar cluster $N_{709}TS_{711}$ is, aside from the Thr, not directly involved in dimerization, but seems to be responsible for kink-formation, bringing two Phe's (at position 706) in proximity. These observations highlight how small residues may act to change the overall geometry in other ways than simply allowing close packing of the backbones. In the ErbB2-TMD dimer (2N2A), the residues $Ix_3Vx_2LLx_2VLx_2VFx_2L$ make up the helix–helix interface with packing of this highly hydrophobic motif mediated by van der Waals' contacts [61]. In addition, the 2N2A structure also describes dimerization of the C-terminal JM of the ErbB2 through the motif TMx_2LLx_2T [61].

Classical GAS-motifs do not participate in the left-handed group (except for the APP-TMD), where the interaction motifs sometimes can be assigned to heptad motifs. Small residues nonetheless play a role in the left-handed structures, but in the form of the small-x_2-small motif (henceforth termed GAS_{short}) at the helix crossing site, resulting in crossing angles clustering around ~ 20° (Fig. 6). Thus, it appears that when a small-residue motif is part of the interaction motif, the spacing between the small residues dictate topology and hence into which group the homodimer will pack. However, the diversity among the interaction motifs of the left-handed dimers varies more than the right-handed. Even though the small residues Ala and Thr (and Gly) are found in 7/8 of the left-handed structures, they are not always found in pairs of two at the site of helix crossing. This is true for the TMDs of ζζ, FGFR3, DAP12, and APP (2LOH, discussed further below). The ζζ-TMD represents a special case where a disulfide bridge holds the two helices together at the N terminus. The FGFR3-TMD helices cross at a single Gly, but no other small residues are found in the interaction interface (Table 1, Fig. 7). Consequently, the structure has a very distinct curvature, with the N terminus bending away from the overall direction of the helix axis (Fig. 3M). The DAP12-TMD interface is mediated by an Asp residue resulting in an almost parallel configuration that ends up in a left-handed kink (Fig. 3P). For the PDGFR-TMDs, it was found that a long stretch of residues ($Vx_3Ax_2Ax_2VLx_2Ix_3Ix_2M$) make up the interaction surface [64]. Included here is an Ax_2A motif, ensuring close association at the crossing site and left-handed packing. The ErbB3-TMD is the only left-handed structure without any small

Fig. 7. The amino acid sequences of the current SPTMR-TMD structures. Residues of the α-helices are shown in bold face, while hydrophobic side chains are gray, polar are green, positively charged are red, negatively charged are blue. Cys, Gly, and Pro are orange. The sequences have been aligned according to the extent of their α-helices, and for the homodimers, the residues that partake in interhelical interactions are highlighted with a purple box. The CD4, ErbB1 and ErbB2 sequences have for illustrative purposes been cut short (signified by '. . .').

residues in the interaction interface, even though its primary structure harbors Gly, Thr, and Ala (Fig. 7). Instead, the interactions of the ErbB3-TMDs rely on van der Waals' packing and π–π stacking between aromatic side chains and between these and the positively charged side chain of Arg. Although many more structures are indeed needed to fully define the left-handed family of TMD homodimers, the GAS_{short} (small-x_2-small with small being Ala or Gly) motif and stabilization from residues with large side chains seems to account for the more ideal left-handed packings. Interhelix stabilization by aromatic side chains is also common in the structures, occurring in approximately half of all the structures, with a slightly higher frequency in the left-handed group (Table 1).

Currently, the dominating paradigm is that the small-residue sequence motifs dictate the interactions of TMDs, a concept that is not widely applied in the prediction and analysis of oligomerization in soluble proteins [45]. The sequence motif paradigm, while being able to explain some interactions in TMD homodimers, is increasingly being recognized as an oversimplification of the intricate atomic structures and interactions that are not yet fully understood [45,87]. From the structures solved of the mammalian SPTMR-TMDs reviewed here, specific residue motifs may explain predisposition for the packing, whereas the packing does not appear to be predictable from the primary structure. This is probably because the small-residue motifs are only partially responsible for the dimerization site, while 5–10 additional residues partake in stabilizing interactions in the dimer interface. A second reason for the lack of predictability could be the presence of interleaved dimerization motifs of importance for allowing two—or more—helix interaction surfaces, which is discussed further in the following.

Switching requires two interaction surfaces with impact on motif conservation

Some SPTMRs have been shown to act in preformed homodimers whose conformations are changed in response to extracellular ligand binding [18,25,26,96]. This finding has prompted the hypothesis that at least two biologically relevant TMD dimer conformations exist for these SPTMRs; one corresponding to the active receptor state (ON) and one corresponding to the inactive state (OFF). Switching between these states upon activation is anticipated to result in changes in, e.g., the distance between the C and N termini or rotation of the TMD α-helices.

As noted from, e.g., the TMD primary structures of the ErbB family, multiple small-x_2-small or small-x_3-small motifs may be present in a single TMD (Fig. 7). This has stimulated the idea that these motifs each participate in a biologically relevant dimer interface [25,97], but until recently with little experimental support [45]. However, based on cellular studies, chemical cross-linking, NMR data, and MD simulations, it has recently been suggested that ligand-induced activation of the ErbB1 causes switching from an N-terminal to a C-terminal GAS-mediated TMD dimer conformation, leading to separation of the TMD α-helices on the intracellular side [15,17,98,99]. This rearrangement is suggested to release the

intracellular JM region from the membrane bilayer, which instead forms an antiparallel dimer, ultimately triggering kinase activation [15,17,98]. This hypothesis is based on MD simulations by Arkhipov *et al.* [17], investigating the stability of two putative dimer conformations of the ErbB1-TMD formed through two GAS-motifs located N- and C-terminally, respectively. These simulations showed a higher stability of the dimer formed by the Gx_3G motif at the N terminus, consistent with NMR studies of the isolated ErbB1-TMD in bicelles (2M20) [15,17]. In contrast, the C-terminal motif Ax_3G supported a relatively unstable ErbB1-TMD dimer, which was found to be stabilized when other domains of the ErbB1 were present [17]. A very recently solved structure of an ErbB1-TMD C-terminal dimer (2M0B) published by Bocharov *et al.* [62] to some extent support the findings of Arkhipov *et al.*, with the C-terminal dimer utilizing the predicted motif. This second dimer conformation nevertheless deviates substantially from the C-terminal dimer model suggested by Arkhipov *et al.*, and has a smaller N-terminal separation and helix crossing angle [17,62]. Nonetheless, it has been suggested that the ErbB1-TMD dimer structure 2M20 solved by Endres *et al.* [15] represents the ON-state conformation, while the 2M0B structure by Bocharov *et al.* [62] could represent the receptor OFF-state conformation. Bocharov *et al.* [62] further speculated that the two different dimer conformations of the ErbB1-TMD arise due to differences in the membrane mimetics used, with the assumed OFF-state solved in DPC micelles, while the ON-state was solved in bicelles. As micelles are more water permeable than bicelles, it is argued that the more hydrophobic dimerization motif utilized in the OFF-state conformation is favored in a more polar environment, and *vice versa* [62]. Based on this observation, Bocharov *et al.* [62] suggested that the lipid bilayer partakes in transferring ligand binding into TMD switching. However, the primary structures of the two studied ErbB1-TMD variants have important differences that may play a role (Fig. 7). In the 2M20 variant, Met650 and Met668 of the TMD are substituted with Leu and Ile, respectively, the 2M0B variant contains seven additional N-terminal residues compared to the 2M20 variant, while the 2M20 variant contains 24 additional C-terminal residues. These differences may favor and/or disfavor, or even abolish, interactions that are directly or indirectly involved in dimerization, and it should be noted that M650 partakes in the helix–helix interaction of the 2M0B dimer (Table 1). To confirm that the properties of the membrane mimetics indeed favors different dimer conformations, the occurrence of these conformations must be confirmed with the same ErbB1-TMD variant in different membrane mimetics.

Two different dimer conformations have similarly been solved for the ErbB2-TMD by the same group [58,61]. The ErbB2-TMD dimer solved in bicelles (2JWA) appears to utilize a more polar interaction motif ($Tx_3Sx_2VGx_2LVx_2Lx_3F$), while the dimer conformation solved in DPC micelles (2N2A) utilize a more hydrophobic motif ($Mx_2ALx_2LLx_2ALx_2GLx_2R$), in agreement with the findings for the ErbB1-TMD. However, again the primary structure of the studied peptides differs in significant ways; the 2N2A variant harbors 18 additional residues C-terminally compared to the 2JWA variant, while the 2JWA variant harbors three additional residues in the N terminus. It is therefore suggested by the authors that the inclusion of the intracellular JM region in the 2N2A variant, shown to form a parallel homodimer, may favor a different conformation compared to the TMDs in isolation [61]. Although the dimer conformations have not been directly correlated with cellular studies, the authors hypothesize that the 2JWA structure corresponds to the ON-state conformation [58], while the 2N2A structure corresponds to the OFF-state conformation [61]. Surprisingly, the ErbB2-TMD dimer described by the 2N2A structure does not utilize the C-terminal Gx_3G motif, which has been shown to be involved in self-association of the ErbB2-TMD in membrane bilayers by a TOXCAT mutational study [100]. This TOXCAT study successfully identified the Sx_3G motif partaking in the association of the N-terminal 2JWA dimer, and is further supported by a computational study, which found the ErbB2-TMD to homodimerize through either of the two GAS-motifs (N and C termini), and that the TMD could switch through rotation without encountering a prohibitive energy barrier between the two [97]. These conflicting results raise the question of the degree of nativeness of the TMD homodimers in their isolated forms and in membrane-mimicking solvents, and clearly more data are required to uncover fully the mode of signal transduction across the membrane bilayer through the ErbBs. Nevertheless, the finding of two different homodimer conformations of the ErbB1-TMD and ErbB2-TMD are highly interesting and highlight the flexibility and sensitivity of single-pass TMD associations.

Of further note, a second dimer population has been observed in the structural study of the EphA1 homodimer [65]. This second, but sparsely populated dimer conformation appeared upon an increase of the pH from 4.3 to 6.3, resulting in deprotonation of Glu547, and involved a second, C-terminal GAS-motif: Ax_3G (560–564) [65]. Unfortunately, due to the small

occupancy of this presumed second dimer conformation, structure elucidation was not possible [65]. However, based on the observed chemical shift changes, it was proposed that this second dimer conformation consisted of a left-handed dimer with a helix crossing angle of ~ 30° [65], in contrast to the major populated state which was a right-handed dimer. EphA1 has been proposed to function as a pH sensor, and the authors therefore speculate that the switch from a right-handed to a left-handed homodimer structure could be related to EphA1 signal transduction [65]. Acquisition of transfer-NOEs to define the interaction surface of such low-affinity and/or lowly populated dimers is complicated by the low signal-to-noise and/or the limited lifetime of the complex. An alternative way of accessing the structure of such a lowly populated dimer is through the application of relaxation-dispersion-type experiments [101]. These experiments have been applied successfully to folded proteins [101,102] and recently also to larger membrane proteins such as phospholamban [103] and the ADP/ATP transporter [104]. To the best of our knowledge, relaxation-dispersion-type experiments have never been applied to SPTMR-TMDs, but may harbor great potential in the pursuit of the sequence-encoded dimer reorganization-switch.

Mutations in TMDs that either activate or inactivate SPTMRs in a cellular context provide another source of insight into sequence-encoded switches between different homodimeric conformations. Currently, only two sets of wild-type (WT)–mutant homodimer structures from SPTMRs are available; one set of the APP-TMD [70] and one set of the VEGFR2-TMD [23]. Of these, the structure of the APP-TMD-variant does not relate to receptor signal transduction, leaving only one confirmed and structurally characterized example of both TMD ON/OFF states. In the work on the VEGFR2-TMD, the homodimer of the constitutive active TMD mutant T771E/F778E was found to differ from the WT in three major aspects; enhanced propensity to dimerize, higher flexibility, and a different dimerization interface [23]. The ON-state dimerization interface arose from a 180° rotation of the helices, shifting the dimerization interface to the opposite side of the helices, whereas no difference in the N- and C-terminal gaps was observed [23]. This study implies that the lowest energy dimer of the VEGFR2-TMD is a predimerized OFF-state, and that activation involves rotation to a different dimerization interface representing the ON-state.

Thus, recent studies suggest that the single-pass TMDs are not only capable of forming one specific homodimer and that alternative states may be favored by, e.g., point mutations, interactions of adjacent domains, or membrane mimetics. However, the amount of data that currently exist on homodimer switches remains sparse. Presently, the concept of a sequence-encoded dimer-switch as a player in receptor signal transduction remains not only to be characterized but also to be fully validated at atomic resolution. To solve this challenge, data on the functionality of these receptors *in vivo* must be combined with atomic resolution structural studies. In the following, it will be explored how far structural biology has come in providing high-resolution descriptions of monomeric and homodimeric SPTMR-TMDs.

Structural quality of the TMD structures

Evaluation of the structural statistics

When evaluating a protein structure, two fundamentally different questions are asked: (a) how well does the structure capture the reality of the protein sample, and (b) how well does the protein sample simulate the native structure. The second question is too premature to answer for SPTMR-TMDs given the low number of structures, but has been reviewed extensively elsewhere for membrane proteins in general [44,105]. In the following, we will address the first question, and solely review how reliable the SPTMR-TMD NMR structures are in terms of representing the reality of the protein under the experimental conditions.

The quality of NMR structures is usually evaluated on the basis of the root-mean-square deviation (RMSD) of the structural ensemble, the Ramachandran plot quality, the number of experimentally derived structural restraints per residue, and the number of violations of the latter. In the special case of SPTMR-TMD structures, we will focus on the ensemble RMSD of all the heavy atoms in the structured region and the number of restraints per residue (Table 2). Factors that have not been included in the evaluation, such as Ramachandran plot statistics, restraints violations, and ProSA2 *z*-scores [106] are overall very similar for all the structures and their values acceptable. Furthermore, dihedral angle restraints that have been predicted by TALOS based on observed chemical shifts have been applied in all the structure calculations, but have not been included in Table 2 because the relative number of predicted dihedral angle restraints applied is similar for all the structures.

Table 2. Structural statistics of available structures of WT SPTMR-TMDs.

PDB code	Abbrev.	Non-native residues	Solvent	Medium-range NOEs	Long-range NOEs	Intermonomeric NOEs	H-bond restraints	Other	RMSD, ordered region, all heavy (Å)	Restraints per residue	ProSA *z*-score	Ref.
Monomers												
2KLU	CD4	N_t-GPLVPRGS	DPC	178	0	–	N		6.1 13.6	6.85	0.07	[54]
2MXB	mEPOR	-6xH-C_t	DPC	53	0	–	Y		1.48 ± 0.23	1.51	− 1.39	[53]
2MV6	EPOR	-6xH-C_t	DPC	42	0	–	Y		1.12 ± 0.22	1.20	− 1.23	[52]
2N71	PRLR	N_t-GS-	DHPC	142	0	–	Y	$^3J(H^N\text{-}H^\alpha)$ couplings	0.40 ± 0.09	5.68	− 0.09	[20]
2MFR	IR	-E-6xH-C_t	DPC	22	0	–	Y		1.12 ± 0.20	0.76	1.26	[49]
1IIJ	mErbB2	0	TFE	N/A	N/A	–	N		N/A	N/A	0.20	[51]
2L8S	Int. αI	-6xH-C_t	LDAO	61	0	–	N	32 RDCs	1.41	3.32	− 0.76	[46]
2RMZ	Int. β3	0	Bicelle	0	0	–	N	153 RDCs	0.66*	5.10	− 0.50	[47]
2RN0	Int. β3	0	DPC	0	0	–	N	90 RDCs	0.81*	3.00	− 0.36	[47]
2K1A	Int. αIIβ	N_t-G-	Bicelle	78	12	–	N	83 RDCs	0.8*	7.21	0.23	[48]
2LLM	APP	N_t-GS-	DPC	75	0	–	Y		1.59 ± 0.28 0.55 ± 0.10	2.68	0.64	[55]
2N5S	ErbB1	N_t-GS--GG-C_t	DPC	81	0	–	Y		0.86 ± 0.31 1.70 ± 0.22	3.12	− 0.51	[50]
2LP1	APP	-QGRILQISITL AAALE6xH-C_t	LMPG	22	0	–	Y	63 PREs 26 RDCs	0.73 0.51	4.44	1.41	[56]
Dimers												
2M20	ErbB1	-KLWS-C_t	Bicelle	N/A	N/A	19	Y	MD	0.76 1.08	N/A	− 0.65 0.65	[15]
2M0B	ErbB1	0	DPC	128	20	20	Y		0.98 ± 0.31	3.08	− 0.68 − 0.67	[62]
2JWA	ErbB2	0	Bicelle	106	20	20	Y	MD	0.92 ± 0.17	2.10	1.31 1.32	[58]
2N2A	ErbB2	-G-C_t	DPC	306	28	28	Y	MD	1.18 1.86	5.22	0.39 0.51	[61]
2L9U	ErbB3	N_t-M--R6xH-C_t	DPC	181	20	20	Y		0.91 ± 0.14	3.05	− 0.23 − 0.41	[59]
2LCX	ErbB4	0	Bicelle	152	16	16	Y		0.83 ± 0.17	2.90	1.68 1.8	[60]
2LZL	FGFR3	0	DPC/SDS	160	26	26	Y	Two side chain to backbone H-bonds	0.73 ± 0.18	3.72	1.15 1.05	[63]
2L6W	PDGFR	N_t-G-	DPC	104	46	46	Y	MD, solid state, OCD	0.67 ± 0.08	2.59	0.18 0.02	[64]
2M59	VEGFR2	0	DPC	164	18	18	Y	MD	0.74 ± 0.12	2.53	− 0.49 − 0.48	[23]
2K1L	EphA1	0	Bicelle	164	28	28	Y	MD	0.64 ± 0.19	3.43	− 1.11 − 0.79	[65]
2K9Y	EphA2	0	Bicelle	122	48	48	Y	MD	0.86 ± 0.24	3.70	1.05 0.46	[66]
2LZ3	APP	0	DPC	280	48	18	Y	88 RDCs	1.28 ± 0.34	7.43	1.08 1.10	[70]
2LOH	APP	N_t-GS-	DPC	308	14	12	Y	Two side chain to backbone H-bonds	1.48 ± 0.22 1.02 ± 0.20	5.55	0.97 0.86	[71]
2L34	DAP12	N/A	N/A	N/A	N/A	26	N/A	Disulfide bridge	0.75	N/A	0.02 0.13	[67]

Table 2. (Continued).

PDB code	Abbrev.	Non-native residues	Solvent	Medium-range NOEs	Long-range NOEs	Intermonomeric NOEs	H-bond restraints	Other	RMSD, ordered region, all heavy (Å)	Restraints per residue	ProSA *z*-score	Ref.
2KPE	GpA	N_t-R--RRL-C_t	DPC	88	34	34	Y	MD	0.94 ± 018	2.26	0.70 0.60	[72]
2KPF	GpA	N_t-R--RRL-C_t	Bicelle	86	28	28	Y	MD	1.07 ± 0.15	2.20	1.29 1.11	[72]
1AFO	GpA	0	DPC	N/A	N/A	N/A	N/A	J-couplings	0.75	N/A	1.07 0.77	[33]
2HAC	ζζ	0	5 : 1 DPC : SDS	N/A	N/A	46	N	Disulfide bridge, 35 RDCs	0.65	N/A	0.74 0.56	[68]
2MK9	TLR3	N_t-M-	DPC	166	16	16	Y		1.07 ± 0.18	3.03	− 1.03 − 1.05	[69]

Except for the ProSA *z*-scores, the information of this table has been obtained from the published paper of each structure. Abbrev.: short abbreviated name for the receptor of the TMD. *H*-bond restraints: 'Y' if hydrogen bond restraints were employed in the structure calculation, 'N' if not. Other: other structural restraints or refinements. MD is molecular dynamics simulations, while OCD is oriented circular dichroism. RMSD, ordered region, all heavy: RMSD of the ensemble for all heavy atoms of the ordered region. '*' highlights structures where the region of the provided RMSD were not clearly defined. Restraints per residue: number of medium- and long-range NOEs, PREs, and RDCs divided by number of residues in the ordered region of the TMD (it is assumed that the contribution from flexible regions to the total restraint count as defined here is negligible). As a point of reference on how many restraints per residue that may be expected to be collected for each TMD structure, we used the same definition to calculate the number of structural restraints per residue for three well-defined, soluble, four-α-helix bundle proteins solved by solution-state NMR spectroscopy; the structures of prolactin [108], ACBP [109], and AML1-ETO [110] (see main text). ProSA *z*-scores [106]: the ProSA *z*-score measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations [106]. *Z*-scores outside a range characteristic for native proteins indicate erroneous structures. *Z*-scores were calculated for each monomer, and all structures fell within the expected range for structures of their size. Ref.: reference describing the structure determination.

The RMSD of a structural ensemble reports on the precision, but not the accuracy, of the derived structures. A well-defined NMR structure of a small protein should have an RMSD of the heavy atoms in the structured region below 1.0 Å [107]. Overall, the RMSDs are reasonable for most of the TMD structures, but one structure stands out with a reported RMSD of 6.1 Å for the TMD helix and 13.5 Å for a minor helix (CD4). However, this high RMSD improves by aligning only the TMD helix. For the remaining structures, the average RMSD is 1.0 Å for the monomers, while it is 0.9 Å for the homodimers. However, not reflected in the RMSDs, but observed when the ensembles are visually inspected, is a scissor-like deviation in some of the dimer structures. This lack of convergence is apparent from significant deviations in the distances between the N terminus and the C terminus, respectively, of the α-helices. In particular, the GpA-TMD structure (1AFO) has an N- and C-terminal deviation of 2 and 10 Å, respectively, and the APP-TMD structure (2LZ3) and the TLR3-TMD structure have N- and C-terminal deviations of 6 and 7 Å, respectively. Even though these structures pinpoint the dimerization interfaces, they provide a very low resolution of the relative positions of the helix N and C termini, which are critically important for interpreting the TMD structure in a full-receptor context.

Evaluating the accuracy of an NMR-derived structure is not straightforward. However, the number of experimentally derived structural restraints per residue is an important determinant. A high number of structural restraints per residue is necessary to define the structure but does not in itself guarantee high structural reliability, whereas a low number should be regarded as problematic. Here, we define the number of structural restraints per residue as the sum of medium-range ($1 < |i - j| < 4$) NOEs, long-range/intermonomeric NOEs, paramagnetic relaxation enhancements (PREs), and residual dipolar couplings (RDCs) divided by the number of residues in the TMD α-helix (Table 2). Predicted dihedral angles are considered indirect restraints, and we therefore found it more transparent to not include them in the restraint counts. In this regard, we note that when there is a large number of NOE-derived distance restraints, chemical shift-derived restraints play only a smaller role in refining the structure, whereas in the case of few distance restraints, the structure is to a large

extent supported by the dihedral angle restraints. The fewer long-range contacts inherent to single-pass TMD structures are expected to result in fewer NOEs per residue than for helices in globular proteins. For comparison, three well-defined, soluble, four-α-helix bundle proteins solved with solution-state NMR spectroscopy [108–110] were found to have 10.6–11.1 structural restraints per residue in ordered regions, which was reduced to 5.8–6.6 restraints per residue (average of 6.1) if long-range NOEs were excluded (Table 2). With this in mind, the SPTMR-TMD structures can be divided into three groups; structures with <2.6 restraints per residue [10 standard deviations (STDs) below the average of the three four-α-helix bundles], structures with 2.6–5.0 restraints per residue (between 10 and 3 STDs below average), and structures with more than 5.0 restraints per residue. In four cases, a per residue dataset was not available [TMD structures of ErbB1, GpA (1AFO), ζζ, DAP12]. Of the remainder, 8/27 structures have <2.6 restraints per residue, 12/27 have between 2.6 and 5.0 restraints per residues, while 7/27 have more than 5.0 restraints per residue (Table 2). It should be noted that several of the structures are additionally supported by molecular dynamics (MD) simulations or data collected with other techniques (Table 2).

As the structures of the TMDs are derived from NMR-restraints through restrained molecular dynamics simulations [107], a low number of structural restraints essentially means that the structure determination is more similar to an unrestrained molecular dynamics simulation (albeit currently with relatively simple force fields). As an example, the IR-TMD structure has with 0.76 structural restraints per residue the lowest number of all the structures. The ensemble has a relatively high precision (RMSD = 1.1 Å); however, when the ensemble is inspected visually, it is clear that especially the side chains have low resolution. At the other end of the scale, the APP-TMD structure 2LZ3 has the highest number (7.43) of structural restraints of all the structures. However, the RMSD is higher than the average (1.3 Å), and the homodimer ensemble has N- and C-terminal deviations of 6 and 7 Å, respectively, although the relative orientations of the side chains are well defined. These two examples illustrate how evaluation of the reliability of NMR-derived structures is not straightforward and requires evaluation of several factors, and highlights that for these structures no strong correlation is apparent between the RMSD of the structural ensembles and the number of applied restraints (Table 2). Overall, the current monomeric and homodimeric single-pass TMD structures appear to have fewer experimental restraints per residue than found for the NMR structures of soluble, globular proteins.

Use of hydrogen bond restraints

Hydrogen bond restraints have not been included in the restraint counts above, but are commonly used (in 23 out of the 30 structures for which this information is available). Introduction of backbone hydrogen bonds restrain the backbone structure significantly and should therefore be applied with caution. Due to the limited amount of tertiary structure in single-pass TMD structures, these restraints may act as an aid in defining the curvature of the helices, which, however, is only truly definable through acquisition of a sufficient set of RDCs. In some cases, the use of the hydrogen bond restraints, and therein the assumption of a relatively straight α-helix, appears to be based loosely on experimental data. In most studies, backbone amide hydrogen-to-deuterium (H-D) exchange experiments are used as justification for the inclusion of hydrogen bond restraints [49,50,52,53,56,58,60,62,63,70,72], while the origin of the assignment in other studies is unclear [15,23,59,61,64,65,69]. As a rule of thumb, amides engaged in hydrogen bonds exchange slowly, while other amides exchange fast [111,112]. However, H-D exchange rates also depend on the degree of burial of an amide [111,113]. A study on the effects of detergent on hydrogen exchange rates of model peptides concluded that the hydrophobicity found in the core of micelles slowed down hydrogen exchange rates with a similar magnitude as hydrogen bonding, an effect contributed to water exclusion by the hydrophobic interior [114]. H-D exchange experiments of membrane proteins are typically performed by transferring the TMD-micelle complex from H_2O to D_2O and evaluating which amide protons undergo exchange with deuterium by analysis of peak intensities in ^{1}H-^{15}N-HSQC spectra (a methodology described in [115]). In this way, amides that do not exchange are identified as hydrogen bond donors. However, in light of the above considerations, it is problematic to use H-D exchange data as a sole argument for the introduction of hydrogen bond restraints in proteins embedded in membrane mimetics. Residues in a kinked or curved helix, but shielded by the micellar structure, may incorrectly be interpreted as being engaged in a hydrogen bond, thus leading to a straight α-helix in the resulting structural ensemble. For these reasons, caution should be taken in the interpretation of H-D exchange data on membrane proteins, and these

should not be used as a sole argument for introduction of hydrogen bond restraints. If hydrogen bond restraints are used, they should be introduced in a later iteration of the structure calculation, and generally only when the structure is otherwise significantly supported by, e.g., NOEs or coupling constants [116].

Examples where hydrogen bond restraints may potentially have overly constrained the structures are the monomeric structures of the mouse and human EPOR-TMDs (2MXB, 2MV6) [52,53] (Fig. 2). As mentioned, these structures consist of one straight α-helix that does not only comprise the predicted TMD region but also continues 10 residues into their JM regions without any apparent flexibility [52,53]. However, no NOEs have been assigned for H249-L253 of the hEPOR-TMD or around H249 of the mEPOR-TMD, backbone relaxation data for both EPOR-TMDs revealed increased internal dynamics in the JM region, and the H-D exchange experiments only supported slow exchange of the backbone amides until L246 [52,53]. Furthermore, when the secondary chemical shifts (SCSs) are calculated based on the deposited C^{α} and H^{α} chemical shifts and predicted C^{α} and H^{α} random coil chemical shifts from Refs. [117,118], it is evident that a lower propensity for α-helical structure is found in the JM region. As the currently available predicted random coil chemical shifts are based on data collected in isotropic aqueous solvents, exact estimation of the α-helix population in the EPOR-JM is precluded by the presence of detergents. However, the relative differences in the average SCSs suggest an ~ 50% reduction of the α-helix population in the JM regions compared to the TMDs. Thus, no direct experimental evidence appears to support that the fully formed TMD α-helix extend into the JM region, but due to an overall low amount of restraints (Table 2), excessive restraining may not cause violations in this case. Furthermore, because these regions appear to consist of a mixture of helical and nonhelical conformations, ensemble refinement, or other ensemble modeling methods may be required to determine such flexible conformations.

Low resolution suggested by doubly determined structures

As mentioned above, inconsistencies between doubly determined structures are present in the collection. The structures of five different receptor TMDs have been solved by different groups: the homodimer of the GpA-TMD (2KPE/2KPF vs. 1AFO), the monomer of the APP-TMD (2LLM vs. 2LP1), the homodimer of the APP-TMD (2LZ3 vs. 2LOH), the homodimer/monomer of the ErbB1-TMD (2M0B vs. 2M20 vs. 3N5S), and the homodimer of the ErbB2-TMD (2N2A vs. 2JWA). Of these, 2M0B vs. 2M20 and 2N2A vs. 2JWA represent unique dimer structures, discussed above, while the similar structures will be discussed below.

The GpA-TMD homodimers, solved in DPC micelles and DMPC bicelles by two different groups (1AFO in DPC micelles [33], and 2KPE and 2KPF in DPC micelles and DMPC bicelles, respectively [72]) have very similar backbone structures. Only a small deviation in the last helical turn is immediately apparent, giving a variation in the helical C-terminal distances of 2.5 Å between the structures solved in DPC. However, when the side-chain conformations are compared, numerous differences are found. The RMSD of the heavy atoms in the helical region between the two structural ensembles solved in DPC is 3.1 Å (1AFO vs. 2KPE), while, surprisingly, the RMSD of the heavy atoms in the helical region between the 1AFO solved in DPC micelles and 2KPF solved in bicelles is only 2.9 Å (solved by two different groups). It thus appears that the observed differences are not related to differences in the applied membrane mimetics. For comparison, the same region has an RMSD of all heavy atoms in the helical region of 2.8 Å between the GpA in DPC (2KPE) and in bicelles (2KPF) when solved by the same group. Both 1AFO and 2KPE/2KPF are primarily supported by NOEs (Table 2), but side-chain dihedral restraints derived from quantitative J-couplings have additionally been applied in the structure calculation of 1AFO [33]. Upon visual inspection of the structural ensembles, the accuracy of the side-chain conformations appears higher in the 2KPE/2KPF structures and Ile88 (shown to make interhelical interactions in the 2KPE/2KPF structures) is slightly turned away from the dimerization interface in the 1AFO structure. It should nonetheless be noted that 1AFO is a pioneering structure of the field solved 14 years earlier than the 2KPE/2KPF structures.

The monomeric structures of the APP-TMD solved by two different groups in two different detergents (2LLM in DPC [55] and 2LP1 in LMPG [56]), are clearly different. The RMSD of the heavy atoms in the major helix region of the two structures is 4.2 Å, and compared to 2LLM, the 2LP1 structure has a major kink near Gly708 and Gly709. The presence of this kink was confirmed by EPR double electron–electron resonance experiments on the APP-TMD in lipid vesicles [56]. None of the structures have a high number of restraints per residues, and the 2LP1 structure has no reported NOEs between side chains [56].

Two different groups have solved the structure of the homodimeric APP-TMD in DPC micelles (2LZ3 [70] and 2LOH [71]), with the 2LZ3 structure being a right-handed dimer and the 2LOH a left-handed dimer. With a helix crossing angle of − 25°, 2LZ3 is along with 2N2A and 2M0B part of an outlier group of the right-handed group II, denoted II* (Fig. 6). The helix crossing angle of the left-handed structure 2LOH on the other hand fits well within the left-handed group III with a crossing angle of 23° (Fig. 6). It is also evident that the curvature of the two structures is slightly different, though exact comparison of the two structures is complicated by the different handedness of the helix–helix packing. Even though the two structures employ almost the same dimerization motif (Table 1), slight differences are also observed here. Val710 and Val717 participate in the interhelical interactions in the 2LZ3 structure but not in 2LOH, while Leu723 forms interhelical interactions in the 2LOH structure, but not in 2LZ3.

The homodimeric structure of the ErbB1-TMD including the JM region has been solved in DMPC bicelles by Endres *et al.* [15]. Recently, a paper by Arseniev *et al.* [50], describing the structure of monomeric ErbB1-TMD-JM in DPC micelles, also presented results on a dimeric ErbB1-TMD-JM structure in DMPC bicelles in apparent conflict with those presented by Endres *et al.* [15,50]. As argued by Arseniev *et al.* [50], the differences essentially boil down to the validation of dimer formation under the applied conditions and discrepancies in the data collected on the JM region by the two groups [50]. However, and as noted above, the ErbB1-TMD-JM variant studied by Endres *et al.* [15] harbors the substitutions M650L and M668I, is seven residues longer, and contains an additional KLWS sequence at its C terminus compared to that studied by Arseniev *et al.* [50]. Thus, this may effectively change the molecular properties of the ErbB1-TMD and hence be the origin of the differences. If so, this further supports that the properties of SPTMR-TMDs are highly sensitive to even small changes in sequence, also when these are far from the dimer interaction site. This may be a consequence of additional lipid/detergent–protein– or protein–protein interactions leading to allosteric changes.

In summary, it is concerning that all the double-determined structures have discrepancies bigger than expected from their accuracies, which raise further concern regarding quality, reproducibility, or influence by changes in construct-length and/or experimental conditions. Due to the essential roles of the SPTMR-TMDs in the very delicate signal transduction of their receptor, high-resolution structures also at various conditions are required if a more profound understanding of these mechanisms are to be obtained.

Utilizing the full NMR toolbox

As evident from Table 2 and outlined above, only a modest number of NOE restraints have been collected on the currently solved SPTMR-TMD structures as a group. The reason for the poor output of NOEs from single-pass TMDs is probably multifold, with increased linewidth due to the large-sized micelle–protein complexes, few protein–protein contacts, and transient interactions as the suspected main causes [40,119,120]. With the inadequacy of the usual primary NMR-work horse, the NOE, additional restraints, as well as new or optimized NMR methods are evidently needed to increase the qualities of the structures of these proteins. One possibility is to increase the use of RDCs, which are unique in being global restraints. An adequate set of RDCs in principle allows for the determination of the global relative orientation of atomic bonds [121], which is particularly useful for SPTMR-TMD structures for defining the α-helical curvature. The downside of this method nonetheless is the need for partial alignment of the protein, which can be of significant technical challenge for membrane proteins in particular. For the six SPTMR-TMDs structures solved with the aid of RDCs (Table 2), weak alignment of the micelle-embedded TMDs was accomplished using strain-induced alignment in a polyacrylamide gel [46–48,56,68,70], in addition to G-tetrad DNA for the APP-TMD structure 2LZ3 [70].

Another option already available in the NMR toolbox is the measurement of long-range distances using PREs induced by paramagnetic spin labels. The introduction of a paramagnetic center in a TMD enhances the relaxation process of nearby nuclei and thus lowers the intensity of the NMR signals in its surroundings in a distance-dependent manner, which may be converted to distance restraints for structure calculations [122]. PRE restraints cover longer distances than NOEs (10–25 Å), but are less precise [123]. However, a systematic study has shown that PRE-derived distance restraints may provide sufficient data to obtain a backbone structure of α-helical membrane proteins with an accuracy of 1.5 Å (backbone RMSD) [119]. PREs have so far only been utilized to collect distance restraints for one SPTMR-TMD, the monomeric APP-TMD structure 2LP1 [56].

Hence, as an alternative to collecting more comprehensive sets of NOEs, we suggest that improved accuracy and precision of SPTMR-TMD structures may be obtained by measuring both RDCs and PREs. Furthermore, MD simulations of the experimentally

determined structures are often performed to support the experimental data [15,23,58,63–66,72], which with the continuous improvement of force fields may be of even further value in the future [124]. In this regard, it should also be noted that NMR structure calculations in general involve MD simulations with the experimental data as restraints. These are currently performed in vacuum, typically with the option of additional refinement in water. It is unclear how these approximations affect TMD structures, and future efforts to allow refinement in a micelle or lipid bilayer may be rewarding in terms of improving the precision. We note, however, that to make this possible, we need to have a better understanding of the structural properties of the membrane mimetics used, and how the SPTMR-TMDs are embedded within them. Finally, increased access to affordable fully deuterated detergents and lipids could allow for more groups to engage in the endeavor as well as increase the quality of the ^{13}C-detected NOESY spectra considerably.

Are we missing out on the dynamic oligomeric state? How to study a symmetric and flexible dimer?

Gaining control of a weak dimer

The low number of currently available homodimeric TMD structures from SPTMRs is likely a reflection of the challenges associated with structural studies of these symmetric and flexible dimers. A high conformational flexibility and modest affinity of the TMD homodimers are presumably a prerequisite for their anticipated ability to switch between conformations; these properties however also clearly challenge structural studies of SPTMR-TMD dimers. As discussed, the overall quality of the available structures appears limited, and, in addition, the oligomeric state of the protein is not always clearly defined. The homodimeric state of most of the TMDs has mainly been confirmed under the applied conditions by acquisition of interhelical NOEs [15,23,58–63,65,66,69,71,72], sometimes in combination with cross-linking experiments [64] or with the detection of size changes as estimated from backbone relaxation rates [61,62,69] or ultracentrifugation [70]. Interhelical NOEs obtained in filter-based experiments should preferably be compared to background reference spectra of the protein in the absence of unlabeled protein to avoid misinterpretation of strong intramolecular peaks leaking through the filter due to imperfect isotope labeling [50], but it is rarely clear if these have been acquired. In the structural studies of the monomeric TMDs, it is also not always shown that the TMD is in fact monomeric under the applied conditions. This lack of evaluation of the oligomeric state is probably a result of the complications introduced by their embedment in membrane mimetics, which renders classical size estimations dependent on two unknowns (the extent of oligomerization of the protein and the size of the membrane mimetics). Detection of interhelical NOEs is, when used properly, a reliable measure of a significant population of dimer. Nonetheless, interspecies NOEs are resource-heavy measurements that are not suitable for evaluating large sets of conditions and are not easily, if at all possible, acquired on transiently interacting proteins [120]. Furthermore, they are not appropriate to confirm the absence of oligomerization for studies of the monomeric state. A possible strategy is utilization of paramagnetic tags [48], the implementation of which nonetheless may require prior knowledge of the dimerization site, the generation of a series of mutants, as well as additional technical challenges associated with the use of hydrophobic labels in the context of a membrane-mimicking environment. Thus, development of high-throughput methods for reliable determination of the oligomeric state of micelle-embedded TMDs under a large set of NMR-suitable conditions is essential for the development of the field. One promising lead is the recent development of native mass spectrometry on membrane proteins, allowing acquisition of spectra on intact membrane protein complexes embedded in micelles [125]. These methods are, however, still not sufficiently quantitative to determine the extent of oligomerization.

A particular challenge is that of stabilizing membrane protein oligomers in non-native membrane mimetics. It is becoming increasingly clear that the lipid bilayer may participate in TMD associations through sequence independent effects such as membrane thickness and charge [126,127]. Furthermore, sequence dependent effects mediating specific lipid binding to TMD helices have been suggested to play important roles in the regulation of the monomer–dimer equilibrium, both by counteracting and enhancing dimerization [128]. These effects are naturally not well simulated in the membrane mimetics that allow structural studies by solution-state NMR spectroscopy, putatively destabilizing the TMD homodimers further. Several studies have reported that the monomer–dimer equilibrium for some TMDs may be manipulated, and thereby studied, through changes in the detergent-to-protein (DP) or lipid-to-protein (LP) ratio, as first reported by Fisher *et al.* [129] on the GpA-TMD. This is essentially analogous to the common approach of diluting a protein complex with the purpose of pushing

the equilibrium toward the monomer conformation to allow its dissociation constant to be determined. In addition to GpA, monomer–dimer transitions as a result of changes in the DP ratio have been reported for many of the TMDs of which structures have been reported, e.g., the homodimers of FGFR3, VEGFR, ErbB1-4, APP, and TLR3 [23,59–63,69,71,130].

Solution-state NMR spectroscopy is a powerful tool for studying monomer–dimer transitions as a consequence of changes in e.g., DP ratio, offering detection and residue-level mapping of the formation and dissociation of low to moderate affinity complexes, such as e.g., homodimers, as well as providing structural information. Even in cases where multiple oligomeric states occur simultaneously, data may under certain conditions be extracted and assigned to each state. Methodologies for utilizing solution-state NMR spectroscopy for investigation of TMD oligomerization based on manipulating the DP or LP ratio have been developed [130,131]; however, a high-throughput method for confirming that the chemical shift changes observed stem from changes in oligomeric states is still desirable. The major drawback is additionally that the detergent concentrations applicable for NMR studies have both upper and lower limits. The upper limit is caused by an increase in sample viscosity, and thereby a slowdown in molecular tumbling, leading to line broadening [132], with the exact concentration limit dependent on the nature of the detergent and the experimental conditions. The lower limit arises because it is difficult to properly analyze data obtained with detergent concentrations below the critical micelle concentration (CMC) of the detergent, as the nature of protein/detergent micelles under these conditions is poorly understood. Consequently, only systems undergoing complete monomer–dimer transitions within the applicable detergent range are interpretable. Furthermore, with the exception of APP-TMD in LMPG micelles [130], the majority of TMD-detergent systems studied so far exhibit slow exchange on the NMR timescale of the monomer–dimer transition. Consequently, NMR methodologies for quantitatively studying systems in fast exchange are still lacking behind.

Solution-state NMR studies of the TMD monomer–dimer transition

Determining the kinetics of monomer–dimer transitions in detergent solvent is a rather complex case, in which factors imposed by the solvent system, such as the CMC, properties and participation of the detergent, as well as micellar collisions are possible contributors to the observed behavior. Several models have been proposed to aid the derivation of formalisms to describe the oligomerization of single-pass TMDs in such systems: the continuous solvent model [133,134], the detergent-release model [129,134], and the micellar solvent model [131]. In essence, the continuous solvent model assumes that the detergent acts as a solvent in the dimerization process, i.e., that dimerization in micellar solvent behaves similarly to dimerization in water or a lipid bilayer. Such an assumption is mainly applicable when the monomer–dimer transition occurs within the same micelle (e.g., at low DP ratio). In the detergent-release model, the detergent is not treated as a solvent but rather as a participant in the dimerization process, a description found best applicable at a high DP ratio and to relatively strong dimers. Lastly, the micellar solvent model is based on the assumption that dimerization and dissociation occur only upon collision and decay of the micelles, respectively. The associated formalism only applies when the micelle collisions are frequent and the dimer dissociation rates are slow on the NMR timescale, and transitions within a single micelle therefore become negligible. A detailed description of all the models and formalisms for describing the equilibrium can be found in Ref. [131]. That no single formalism or model is capable of describing all monomer–dimer transitions highlights their complexity, and that much remains to be understood. The homodimerization of the APP-TMD has for example been studied in both DPC micelles [71] and in LMPG micelles [130], where the monomer–dimer is in slow exchange on the NMR timescale in DPC, but in fast exchange in LMPG. That is, the monomer–dimer transition of the APP-TMD fits within the detergent-release model and the continuous solvent model, depending on the detergent rather than the intrinsic nature of the TMD association. This is in line with the observation that the magnitude of the dimer dissociation constant is dependent on the applied detergent [129,134], and thus the question of the biological relevance of such estimations of the equilibrium kinetics may be raised. Even when these measurements are conducted in lipid bilayers, the strength of the associations varies considerably with membrane composition [126,127], a scenario which might reflect biological relevance. It can be argued, as suggested by Mineev *et al.* [131], that these estimations may be used to compare the strength of dimer formation of TMDs measured in the same membrane-mimicking environment. However, as mentioned, different detergents and lipids do not affect structure, stability, and functionality of membrane proteins similarly and

this interplay is currently not well understood. The question is therefore if the same membrane mimetic environment also affects different TMD monomer–dimer equilibria differently, and the measured equilibrium kinetics thus becomes an equation with too many unknowns to provide any biologically meaningful results. It should be noted that most of these studies have been conducted on the GpA-TMD only, which is generally considered to form a very stable homodimer [135], presumable also more tolerant to varying conditions.

In terms of atomic resolution structural knowledge on the monomer–dimer transition, only the TMDs of ErbB1, ErbB2, and APP are represented by structures of both a monomer and a homodimer, and in case of ErbB2, the monomer and dimer does not originate from the same species. Additionally, as discussed, controversy around the structures of the ErbB1-TMD exists, while the characteristics of the APP-TMD are unclear due to the multiple structures representing the monomeric and dimeric APP-TMD. Consequently, little direct knowledge is available on the details of the structural transformation of the TMDs upon monomer to dimer transition. The handful of monomeric, helical TMDs, however, clearly demonstrates that the formation of secondary structure is uncoupled from oligomerization. It should, however, be noted that the ErbB1-TMD α-helix has been found to be extended upon homodimerization [62].

When biology awaits biophysics

From our analyses of the contemporary structures of SPTMR-TMDs, we were surprised to discover that over a period of almost 20 years since the first structure was reported [33], only a total of 27 unique structures are now available, representing 21 receptors. Given the abundance of these proteins in the eukaryotic proteome with more than 1300 different SPTMRs in humans alone, this means that <2% currently act as structural representatives of the entire family. Thus, structural biology is critically lacking behind, and the time where we will be able to extract general properties, where the structures are so many that they will generate predictive power and when we are able to elucidate molecular mechanisms is not even within eyesight. Moreover, as we have highlighted, the quality of the limited number of structures is not always *on par* with expected standards and the situation has resemblance to when the first X-ray structure of myoglobin appeared. This is clearly a problematic situation; the design of TMD-specific pharmaceutical agents, prediction or rationalization of mutational affects in relation to disease, and deciphering of functional mechanisms require high resolution, especially when dealing with a domain carrying so much functionality in so few residues. With the current pace of 1–2 structures solved per year, we have to wait more than 100 years to reach completeness. Adding to this, only the surface has been scratched in terms of evaluating how well the determined structures in membrane mimetics actually represent the reality found in the native membrane bilayer. In this respect, the environmental properties may take on an equally big and convoluted role, and in a cellular context dynamical changes in the lipid composition constantly occur as a result of metabolic changes, which are often close to and even mediated by lipid-binding motifs in the membrane protein itself [128]. Thus, the Anfinsen dogma of how the protein fold is determined by the interplay between the primary structure and its environment must not be neglected.

The question is why these proteins are so problematic and what obstacles are holding the field back? One of the issues is, as discussed, the conformational flexibility in terms of the oligomeric state inherent to these TMDs, and this property challenges the field of structural biology in more ways than merely pushing at the boundaries of the current techniques. The signal transduction model for SPTMRs shown to be dimeric prior to ligand binding is anticipated to entail at least two biologically relevant dimer conformations—the ON- and the OFF-states. However, as structural studies of single-pass TMDs are currently conducted on the TMDs in isolation, detached from the ECD and the ICD, it cannot be directly deduced if the structural state of lowest energy present in the NMR tube is the ON- or OFF-state, or neither. Therefore, validation and interpretation of the structural findings through cellular studies are of paramount importance for the field and coupling between structural biology and cell biology a necessity.

Clearly, plenty of challenges remain to be tackled in the structural studies of these seemingly simple, but intriguingly complex domains, and we have below highlighted a few of the outstanding questions (Box 1). Some issues may be relieved to some extent by other techniques than the currently dominating solution-state NMR spectroscopy entering the scene of SPTMR-TMD structures. Solid-state NMR spectroscopy is emerging as a promising tool for structural studies of membrane proteins, especially because it allows applications of more native-like membrane bilayers [136]. However, solid-state NMR spectroscopy does not have the year-long proficiency in structure determination as solution-state NMR, with

Box 1. Outstanding questions.

- Why are the α-helices of the TMD structures longer than the predicted TM region?
- How do we improve the quality of the NMR structures of SPTMR-TMDs?
- How do we extend structural studies to other SPTMR families?
- What do low-affinity TMD interaction motifs look like, what are their properties and what are the relationships between the linear primary structure and the tertiary and quaternary structure of TMD dimers?
- What is the influence of different membrane mimetics on the structure and dynamics of SPTMR-TMDs, and how do we ensure that we study native structures?
- How is the TMD structures affected by structure calculations in vacuum?
- How can we study SPTMR-TMD monomer–dimer transitions in a biologically meaningful way?
- How do we structurally characterize transient dimers?
- How can different dimer conformations be studied and linked to functionally different receptor states?

only 81 deposited structures in the PDB [32], but is developing fast. Furthermore, this technique requires significantly higher amounts of isotope-labeled protein than solution-state NMR, which nonetheless is increasingly resolved by the development of new efficient production schemes [34,137,138].

Lastly, with the increased resolution obtained with cryo-electron microscopy [139–141], it is possible that within the coming decade, it will be possible to resolve entire SPTMR structures including the ECD and the ICD. This will provide a major breakthrough for structural biology in general and for cellular signaling in particular, but it will not be possible before full receptors can be produced, purified, and stabilized in sufficient amounts. Until then, solution-state NMR spectroscopy may have to carry the torch alone.

Acknowledgements

This work was supported by grants from the Lundbeck foundation (BBK), the Danish Research Council for Health and Disease (BBK, grant number 12-125862) and the Novo Nordisk Foundation (BBK and KLL). We thank the VELUX FOUNDATIONS for a generous grant to purchase new NMR equipment. This is a contribution from the Linderstrøm-Lang Centre for Protein Science. We thank Gitte W. Haxholm for critical reading of the manuscript.

Author contributions

KB wrote the first draft and KB and BBK wrote most of the manuscript with inputs from KLL.

References

1 Wallin E & von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **7**, 1029–1038.
2 Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M *et al.* (1996) Life with 6000 genes. *Science* **274**, 546–567.
3 Overington JP, Al-Lazikani B & Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov* **5**, 993–996.
4 Tusnády GE, Dosztányi Z & Simon I (2004) Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics* **20**, 2964–2972.
5 Pahl MC, Askinazi OL, Hamilton C, Cheng I, Cichewicz K, Kuhn J, Manohar S & Deppmann CD (2013) Signalling via single-pass transmembrane proteins. In eLS, pp. 1–12. John Wiley & Sons Ltd, Chichester, UK.
6 Takada Y, Ye X & Simon S (2007) The integrins. *Genome Biol* **8**, 215.
7 Li E & Hristova K (2006) Role of receptor tyrosine kinase transmembrane domains in cell signaling and human pathologies. *Biochemistry* **45**, 6241–6251.
8 Choudhuri K & van der Merwe PA (2007) Molecular mechanisms involved in T cell receptor triggering. *Semin Immunol* **19**, 255–261.
9 Tayal V & Kalra BS (2008) Cytokines and anti-cytokines as therapeutics—an update. *Eur J Pharmacol* **579**, 1–12.
10 Ramasarma T & Joshi N (2005) Transmembrane domains. In eLS, pp. 1–8. John Wiley & Sons Ltd, Chichester, UK.
11 Zviling M, Kochva U & Arkin IT (2007) How important are transmembrane helices of bitopic membrane proteins? *Biochim Biophys Acta* **1768**, 387–392.
12 De Biasio A, Guarnaccia C, Popovic M, Uversky VN, Pintar A & Pongor S (2008) Prevalence of intrinsic disorder in the intracellular region of human single-pass type I proteins: the case of the notch ligand delta-4. *J Proteome Res* **7**, 2496–2506.

13 Minezaki Y, Homma K & Nishikawa K (2007) Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment. *J Mol Biol* **368**, 902–913.
14 Haxholm GW, Nikolajsen LF, Olsen JG, Fredsted J, Larsen FH, Goffin V, Pedersen SF, Brooks AJ, Waters MJ & Kragelund BB (2015) Intrinsically disordered cytoplasmic domains of two cytokine receptors mediate conserved interactions with membranes. *Biochem J* **468**, 495–506.
15 Endres NF, Das R, Smith AW, Arkhipov A, Kovacs E, Huang Y, Pelton JG, Shan Y, Shaw DE, Wemmer DE *et al.* (2013) Conformational coupling across the plasma membrane in activation of the EGF receptor. *Cell* **152**, 543–556.
16 Hubbard SR & Miller WT (2007) Receptor tyrosine kinases: mechanisms of activation and signaling. *Curr Opin Cell Biol* **19**, 117–123.
17 Arkhipov A, Shan Y, Das R, Endres NF, Eastwood MP, Wemmer DE, Kuriyan J & Shaw DE (2013) Architecture and membrane interactions of the EGF receptor. *Cell* **152**, 557–569.
18 Brooks AJ, Dai W, O'Mara ML, Abankwa D, Chhabra Y, Pelekanos RA, Gardon O, Tunny KA, Blucher KM, Morton CJ *et al.* (2014) Mechanism of activation of protein kinase JAK2 by the growth hormone receptor. *Science* **344**, 1249783–1249783.
19 Matthews EE, Zoonens M & Engelman DM (2006) Dynamic helix interactions in transmembrane signaling. *Cell* **127**, 447–450.
20 Bugge K, Papaleo E, Haxholm GW, Hopper JTS, Robinson CV, Olsen JG, Lindorff-Larsen K & Kragelund BB (2016) A combined computational and structural model of the full-length human prolactin receptor. *Nat Commun* **7**, 11578.
21 Popot JL & Engelman DM (1990) Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29**, 4031–4037.
22 Katragadda M, Alderfer JL & Yeagle PL (2001) Assembly of a polytopic membrane protein structure from the solution structures of overlapping peptide fragments of bacteriorhodopsin. *Biophys J* **81**, 1029–1036.
23 Manni S, Mineev KS, Usmanova D, Lyukmanova EN, Shulepko MA, Kirpichnikov MP, Winter J, Matkovic M, Deupi X, Arseniev AS *et al.* (2014) Structural and functional characterization of alternative transmembrane domain conformations in VEGF receptor 2 activation. *Structure* **22**, 1077–1089.
24 Lau T-L, Kim C, Ginsberg MH & Ulmer TS (2009) The structure of the integrin alphaIIbbeta3 transmembrane complex explains integrin transmembrane signalling. *EMBO J* **28**, 1351–1361.
25 Cymer F & Schneider D (2010) Transmembrane helix–helix interactions involved in ErbB receptor signaling. *Cell Adh Migr* **4**, 299–312.
26 Call ME & Chou JJ (2010) A view into the blind spot: solution NMR provides new insights into signal transduction across the lipid bilayer. *Structure* **18**, 1559–1569.
27 Moore DT, Berger BW & DeGrado WF (2008) Protein–protein interactions in the membrane: sequence, structural, and biological motifs. *Structure* **16**, 991–1001.
28 Li E, You M & Hristova K (2006) FGFR3 dimer stabilization due to a single amino acid pathogenic mutation. *J Mol Biol* **356**, 600–612.
29 Heim EN, Marston JL, Federman RS, Edwards APB, Karabadzhak AG, Petti LM, Engelman DM & DiMaio D (2015) Biologically active LIL proteins built with minimal chemical diversity. *Proc Natl Acad Sci* **112**, E4717–E4725.
30 Lee J, Miyazaki M, Romeo GR & Shoelson SE (2014) Insulin receptor activation with transmembrane domain ligands. *J Biol Chem* **289**, 19769–19777.
31 Bennasroune A (2004) Transmembrane peptides as inhibitors of ErbB receptor signaling. *Mol Biol Cell* **15**, 3464–3474.
32 Berman HM (2000) The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242.
33 MacKenzie KR (1997) A transmembrane helix dimer: structure and implications. *Science* **276**, 131–133.
34 Bocharova OV, Kuzmichev PK, Urban AS, Goncharuk SA, Bocharov EV & Arsenyev AS (2015) Preparation of growth hormone receptor GHR-(254–298) transmembrane fragments in a cell-free expression system for structural studies. *Russ J Bioorganic Chem* **41**, 631–637.
35 Warschawski DE, Arnold AA, Beaugrand M, Gravel A, Chartrand É & Marcotte I (2011) Choosing membrane mimetics for NMR structural studies of transmembrane proteins. *Biochim Biophys Acta* **1808**, 1957–1974.
36 Poget SF & Girvin ME (2007) Solution NMR of membrane proteins in bilayer mimics: small is beautiful, but sometimes bigger is better. *Biochim Biophys Acta* **1768**, 3098–3106.
37 Seddon AM, Curnow P & Booth PJ (2004) Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta* **1666**, 105–117.
38 Bordag N & Keller S (2010) α-Helical transmembrane peptides: a "divide and conquer" approach to membrane proteins. *Chem Phys Lipids* **163**, 1–26.
39 Rath A & Deber CM (2012) Protein structure in membrane domains. *Annu Rev Biophys* **41**, 135–155.
40 Rösner HI & Kragelund BB (2012) Structure and dynamic properties of membrane proteins using NMR. *Compr Physiol* **2**, 1491–1539.

41 Opella SJ & Marassi FM (2004) Structure determination of membrane proteins by NMR spectroscopy. *Chem Rev* **104**, 3587–3606.
42 Nielsen N, Malmendal A & Vosegaard T (2004) Techniques and applications of NMR to membrane proteins. *Mol Membr Biol* **21**, 129–141.
43 Liang B & Tamm LK (2016) NMR as a tool to investigate the structure, dynamics and function of membrane proteins. *Nat Struct Mol Biol* **23**, 468–474.
44 Zhou H-X & Cross TA (2013) Influences of membrane mimetic environments on membrane protein structures. *Annu Rev Biophys* **42**, 361–392.
45 Li E, Wimley WC & Hristova K (2012) Transmembrane helix dimerization: beyond the search for sequence motifs. *Biochim Biophys Acta* **1818**, 183–193.
46 Lai C, Liu X, Tian C & Wu F (2013) Integrin α1 has a long helix, extending from the transmembrane region to the cytoplasmic tail in detergent micelles. *PLoS One* **8**, e62954.
47 Lau TL, Partridge AW, Ginsberg MH & Ulmer TS (2008) Structure of the integrin beta3 transmembrane segment in phospholipid bicelles and detergent micelles. *Biochemistry* **47**, 4008–4016.
48 Lau TL, Dua V & Ulmer TS (2008) Structure of the integrin AlphaIIBeta transmembrane segment. *J Biol Chem* **283**, 16162–16168.
49 Li Q, Wong YL & Kang C (2014) Solution structure of the transmembrane domain of the insulin receptor in detergent micelles. *Biochim Biophys Acta* **1838**, 1313–1321.
50 Mineev KS, Panova SV, Bocharova OV, Bocharov EV & Arseniev AS (2015) The membrane mimetic affects the spatial structure and mobility of EGFR transmembrane and juxtamembrane domains. *Biochemistry* **54**, 6295–6298.
51 Goetz M, Carlotti C, Bontems F & Dufourc EJ (2001) Evidence for an alpha-helix –> pi-bulge helicity modulation for the neu/erbB-2 membrane-spanning segment. A 1H NMR and circular dichroism study. *Biochemistry* **40**, 6534–6540.
52 Li Q, Wong YL, Huang Q & Kang C (2014) Structural insight into the transmembrane domain and the juxtamembrane region of the erythropoietin receptor in micelles. *Biophys J* **107**, 2325–2336.
53 Li Q, Lei Wong Y, Yueqi Lee M, Li Y & Kang C (2015) Solution structure of the transmembrane domain of the mouse erythropoietin receptor in detergent micelles. *Sci Rep* **5**, 1–10.
54 Wittlich M, Thiagarajan P, Koenig BW, Hartmann R & Willbold D (2010) NMR structure of the transmembrane and cytoplasmic domains of human CD4 in micelles. *Biochim Biophys Acta* **1798**, 122–127.
55 Nadezhdin KD, Bocharova OV, Bocharov EV & Arseniev AS (2011) Structural and dynamic study of the transmembrane domain of the amyloid precursor protein. *Acta Naturae* **3**, 69–76.
56 Barrett PJ, Song Y, Van Horn WD, Hustedt EJ, Schafer JM, Hadziselimovic A, Beel AJ & Sanders CR (2012) The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol. *Science* **336**, 1168–1171.
57 The Uniprot Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204–D212.
58 Bocharov EV, Mineev KS, Volynsky PE, Ermolyuk YS, Tkach EN, Sobol AG, Chupin VV, Kirpichnikov MP, Efremov RG & Arseniev AS (2008) Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. *J Biol Chem* **283**, 6950–6956.
59 Mineev KS, Khabibullina NF, Lyukmanova EN, Dolgikh DA, Kirpichnikov MP & Arseniev AS (2011) Spatial structure and dimer–monomer equilibrium of the ErbB3 transmembrane domain in DPC micelles. *Biochim Biophys Acta* **1808**, 2081–2088.
60 Bocharov EV, Mineev KS, Goncharuk MV & Arseniev AS (2012) Structural and thermodynamic insight into the process of "weak" dimerization of the ErbB4 transmembrane domain by solution NMR. *Biochim Biophys Acta* **1818**, 2158–70.
61 Bragin PE, Mineev KS, Bocharova OV, Volynsky PE, Bocharov EV & Arseniev AS (2016) HER2 transmembrane domain dimerization coupled with self-association of membrane-embedded cytoplasmic juxtamembrane regions. *J Mol Biol* **428**, 52–61.
62 Bocharov EV, Lesovoy DM, Pavlov KV, Pustovalova YE, Bocharova OV & Arseniev AS (2016) Alternative packing of EGFR transmembrane domain suggests that protein–lipid interactions underlie signal conduction across membrane. *Biochim Biophys Acta* **1858**, 1254–1261.
63 Bocharov EV, Lesovoy DM, Goncharuk SA, Goncharuk MV, Hristova K & Arseniev AS (2013) Structure of FGFR3 transmembrane domain dimer: implications for signaling and human pathologies. *Structure* **21**, 2087–2093.
64 Muhle-Goll C, Hoffmann S, Afonin S, Grage SL, Polyansky AA, Windisch D, Zeitler M, Bürck J & Ulrich AS (2012) Hydrophobic matching controls the tilt and stability of the dimeric platelet-derived growth factor receptor (PDGFR) α transmembrane segment. *J Biol Chem* **287**, 26178–26186.
65 Bocharov EV, Mayzel ML, Volynsky PE, Goncharuk MV, Ermolyuk YS, Schulga AA, Artemenko EO, Efremov RG & Arseniev AS (2008) Spatial structure

and pH-dependent conformational diversity of dimeric transmembrane domain of the receptor tyrosine kinase EphA1. *J Biol Chem* **283**, 29385–29395.
66 Bocharov EV, Mayzel ML, Volynsky PE, Mineev KS, Tkach EN, Ermolyuk YS, Schulga AA, Efremov RG & Arseniev AS (2010) Left-handed dimer of EphA2 transmembrane domain: helix packing diversity among receptor tyrosine kinases. *Biophys J* **98**, 881–889.
67 Call ME, Wucherpfennig KW & Chou JJ (2010) The structural basis for intramembrane assembly of an activating immunoreceptor complex. *Nat Immunol* **11**, 1023–1029.
68 Call ME, Schnell JR, Xu C, Lutz RA, Chou JJ & Wucherpfennig KW (2006) The structure of the ζζ transmembrane dimer reveals features essential for its assembly with the T cell receptor. *Cell* **127**, 355–368.
69 Mineev KS, Goncharuk SA & Arseniev AS (2014) Toll-like receptor 3 transmembrane domain is able to perform various homotypic interactions: an NMR structural study. *FEBS Lett* **588**, 3802–3807.
70 Chen W, Gamache E, Rosenman DJ, Xie J, Lopez MM, Li Y-M & Wang C (2014) Familial Alzheimer's mutations within APPTM increase Abeta42 production by enhancing accessibility of epsilon-cleavage site. *Nat Commun* **5**, 3037.
71 Nadezhdin KD, Bocharova OV, Bocharov EV & Arseniev AS (2012) Dimeric structure of transmembrane domain of amyloid precursor protein in micellar environment. *FEBS Lett* **586**, 1687–1692.
72 Mineev KS, Bocharov EV, Volynsky PE, Goncharuk MV, Tkach EN, Ermolyuk YS, Schulga AA, Chupin VV, Maslennikov IV, Efremov RG *et al.* (2011) Dimeric structure of the transmembrane domain of glycophorin a in lipidic and detergent environments. *Acta Naturae* **3**, 90–98.
73 Sonnhammer EL, von Heijne G & Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**, 175–182.
74 Eisenberg D, Schwarz E, Komaromy M & Wall R (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* **179**, 125–142.
75 Käll L, Krogh A & Sonnhammer EL (2004) A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **338**, 1027–1036.
76 Kall L, Krogh A & Sonnhammer ELL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* **35**, W429–W432.
77 Jones DT (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* **23**, 538–544.
78 Crick FHC (1953) The packing of α-helices: simple coiled-coils. *Acta Crystallogr* **6**, 689–697.
79 Chothia C, Levitt M & Richardson D (1977) Structure of proteins: packing of alpha-helices and pleated sheets. *Proc Natl Acad Sci USA* **74**, 4130–4134.
80 Mason JM & Arndt KM (2004) Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem* **5**, 170–176.
81 Langosch D & Heringa J (1998) Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils. *Proteins Struct Funct Genet* **31**, 150–159.
82 Teilum K, Olsen JG & Kragelund BB (2015) Globular and disordered—the non-identical twins in protein-protein interactions. *Front Mol Biosci* **2**, 1–6.
83 Jones S & Thornton JM (1996) Principles of protein–protein interactions. *Proc Natl Acad Sci USA* **93**, 13–20.
84 Walters RFS & DeGrado WF (2006) Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci USA* **103**, 13658–13663.
85 Russ WP & Engelman DM (2000) The GxxxG motif: a framework for transmembrane helix–helix association. *J Mol Biol* **296**, 911–919.
86 Senes A, Engel DE & DeGrado WF (2004) Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Curr Opin Struct Biol* **14**, 465–479.
87 Teese MG & Langosch D (2015) Role of GxxxG motifs in transmembrane domain interactions. *Biochemistry* **54**, 5125–5135.
88 Senes A, Gerstein M & Engelman DM (2000) Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with β-branched residues at neighboring positions. *J Mol Biol* **296**, 921–936.
89 Senes A, Ubarretxena-Belandia I & Engelman DM (2001) The C–H***O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc Natl Acad Sci* **98**, 9056–9061.
90 Mueller BK, Subramaniam S & Senes A (2014) A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical Cα-H hydrogen bonds. *Proc Natl Acad Sci USA* **111**, E888–95.
91 Finger C, Escher C & Schneider D (2009) The single transmembrane domains of human receptor tyrosine kinases encode self-interactions. *Sci Signal* **2**, 1–8.
92 Lupas A, Van Dyke M & Stock J (1991) Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164.

93 Gurezka R, Laage R, Brosig B & Langosch D (1999) A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. *J Biol Chem* **274**, 9265–9270.
94 Sal-Man N, Gerber D, Bloch I & Shai Y (2007) Specificity in transmembrane helix–helix interactions mediated by aromatic residues. *J Biol Chem* **282**, 19753–19761.
95 Hong H (2014) Toward understanding driving forces in membrane protein folding. *Arch Biochem Biophys* **564**, 297–313.
96 Tao R-H & Maruyama IN (2008) All EGF(ErbB) receptors have preformed homo- and heterodimeric structures in living cells. *J Cell Sci* **121**, 3207–3217.
97 Fleishman SJ, Schlessinger J & Ben-Tal N (2002) A putative molecular-activation switch in the transmembrane domain of erbB2. *Proc Natl Acad Sci* **99**, 15937–15940.
98 Jura N, Endres NF, Engel K, Deindl S, Das R, Lamers MH, Wemmer DE, Zhang X & Kuriyan J (2009) Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell* **137**, 1293–1307.
99 Lu C, Mi L-Z, Grey MJ, Zhu J, Graef E, Yokoyama S & Springer T (2010) Structural evidence for loose linkage between ligand binding and kinase activation in the epidermal growth factor receptor. *Mol Cell Biol* **30**, 5432–5443.
100 Mendrola JM, Berger MB, King MC & Lemmon MA (2002) The single transmembrane domains of ErbB receptors self-associate in cell membranes. *J Biol Chem* **277**, 4704–4712.
101 Mulder FAA, Mittermaier A, Hon B, Dahlquist FW & Kay LE (2001) Studying excited states of proteins by NMR spectroscopy. *Nat Struct Biol* **8**, 932–935.
102 Teilum K, Smith MH, Schulz E, Christensen LC, Solomentsev G, Oliveberg M & Akke M (2009) Transient structural distortion of metal-free Cu/Zn superoxide dismutase triggers aberrant oligomerization. *Proc Natl Acad Sci* **106**, 18273–18278.
103 Traaseth NJ & Veglia G (2010) Probing excited states and activation energy for the integral membrane protein phospholamban by NMR CPMG relaxation dispersion experiments. *Biochim Biophys Acta* **1798**, 77–81.
104 Brüschweiler S, Yang Q, Run C & Chou JJ (2015) Substrate-modulated ADP/ATP-transporter dynamics revealed by NMR relaxation dispersion. *Nat Struct Mol Biol* **22**, 636–641.
105 Popot J-L & Engelman DM (2016) Membranes do not tell proteins how to fold. *Biochemistry* **55**, 5–18.
106 Wiederstein M & Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* **35**, W407–W410.
107 Kwan AH, Mobli M, Gooley PR, King GF & Mackay JP (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J* **278**, 687–703.
108 Teilum K, Hoch JC, Goffin V, Kinet S, Martial JA & Kragelund BB (2005) Solution structure of human prolactin. *J Mol Biol* **351**, 810–823.
109 Teilum K, Thormann T, Caterer NR, Poulsen HI, Jensen PH, Knudsen J, Kragelund BB & Poulsen FM (2005) Different secondary structure elements as scaffolds for protein folding transition states of two homologous four-helix bundles. *Proteins* **59**, 80–90.
110 Plevin MJ, Zhang J, Guo C, Roeder RG & Ikura M (2006) The acute myeloid leukemia fusion protein AML1-ETO targets E proteins via a paired amphipathic helix-like TBP-associated factor homology domain. *Proc Natl Acad Sci USA* **103**, 10242–10247.
111 Kragelund BB, Knudsen J & Poulsen FM (1995) Local perturbations by ligand binding of hydrogen deuterium exchange kinetics in a four-helix bundle protein, acyl coenzyme A binding protein (ACBP). *J Mol Biol* **250**, 695–706.
112 Teilum K, Kragelund BB & Poulsen FM (2004) Application of hydrogen exchange kinetics to studies of protein folding. In *Protein Folding Handbook* (Buchner J & Kiefhaber T, eds), pp. 634–672. Wiley-VCH Verlag GmbH & Co., KGaA, Weinheim, Germany.
113 Best RB & Vendruscolo M (2006) Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of CI2. *Structure* **14**, 97–106.
114 Spyracopoulos L & O'Neil JDJ (1994) Effect of a hydrophobic environment on the hydrogen exchange kinetics of model amides determined by 1H-NMR spectroscopy. *J Am Chem Soc* **116**, 1395–1402.
115 Veglia G, Zeri AC, Ma C & Opella SJ (2002) Deuterium/hydrogen exchange factors measured by solution nuclear magnetic resonance spectroscopy as indicators of the structure and topology of membrane proteins. *Biophys J* **82**, 2176–2183.
116 Cornilescu G, Hu JS & Bax A (1999) Identification of the hydrogen bonding network in a protein by scalar couplings. *J Am Chem Soc* **121**, 2949–2950.
117 Kjaergaard M, Brander S & Poulsen FM (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J Biomol NMR* **49**, 139–149.

118 Kjaergaard M & Poulsen FM (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR* **50**, 157–165.
119 Gottstein D, Reckel S, Dötsch V & Güntert P (2012) Requirements on paramagnetic relaxation enhancement data for membrane protein structure determination by NMR. *Structure* **20**, 1019–1027.
120 Liu Z, Gong Z, Dong X & Tang C (2015) Transient protein–protein interactions visualized by solution NMR. *Biochim Biophys Acta* **1864**, 115–122.
121 Prestegard JH, Al-Hashimi HM & Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys* **33**, 371–424.
122 Otting G (2010) Protein NMR using paramagnetic ions. *Annu Rev Biophys* **39**, 387–405.
123 Battiste JL & Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355–5365.
124 Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO & Shaw DE (2012) Systematic validation of protein force fields against experimental data. *PLoS One* **7**, e32131.
125 Laganowsky A, Reading E, Hopper JTS & Robinson CV (2013) Mass spectrometry of intact membrane protein complexes. *Nat Protoc* **8**, 639–651.
126 Hong H & Bowie JU (2011) Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *J Am Chem Soc* **133**, 11389–11398.
127 Anbazhagan V & Schneider D (2010) The membrane environment modulates self-association of the human GpA TM domain-Implications for membrane protein folding and transmembrane signaling. *Biochim Biophys Acta* **1798**, 1899–1907.
128 Stangl M & Schneider D (2015) Functional competition within a membrane: lipid recognition vs. transmembrane helix oligomerization. *Biochim Biophys Acta* **1848**, 1886–1896.
129 Fisher LE, Engelman DM & Sturgis JN (1999) Detergents modulate dimerization, but not helicity, of the glycophorin A transmembrane domain. *J Mol Biol* **293**, 639–651.
130 Zhuang T, Jap BK & Sanders CR (2011) Solution NMR approaches for establishing specificity of weak heterodimerization of membrane proteins. *J Am Chem Soc* **133**, 20571–20580.
131 Mineev KS, Lesovoy DM, Usmanova DR, Goncharuk Sa, Shulepko Ma, Lyukmanova EN, Kirpichnikov MP, Bocharov EV & Arseniev AS (2014) NMR-based approach to measure the free energy of transmembrane helix–helix interactions. *Biochim Biophys Acta* **1838**, 164–172.
132 Stanczak P, Horst R, Serrano P & Wüthrich K (2009) NMR characterization of membrane protein-detergent micelle solutions by use of microcoil equipment. *J Am Chem Soc* **131**, 18450–18456.
133 Fleming KG (2002) Standardizing the free energy change of transmembrane helix–helix interactions. *J Mol Biol* **323**, 563–571.
134 Fisher LE, Engelman DM & Sturgis JN (2003) Effect of detergents on the association of the glycophorin a transmembrane helix. *Biophys J* **85**, 3097–3105.
135 Hong H, Blois TM, Cao Z & Bowie JU (2010) Method to measure strong protein–protein interactions in lipid bilayers using a steric trap. *Proc Natl Acad Sci* **107**, 19802–19807.
136 McDermott A (2009) Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR. *Annu Rev Biophys* **38**, 385–403.
137 Bugge K, Steinocher H, Brooks AJ, Lindorff-Larsen K & Kragelund BB (2015) Exploiting hydrophobicity for efficient production of transmembrane helices for structure determination by NMR spectroscopy. *Anal Chem* **87**, 9126–9131.
138 Hu J, Qin H, Li C, Sharma M, Cross TA & Gao FP (2007) Structural biology of transmembrane domains: efficient production and characterization of transmembrane peptides by NMR. *Protein Sci* **16**, 2153–2165.
139 Jiang J, Pentelute BL, Collier RJ & Zhou ZH (2015) Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature* **521**, 545–549.
140 Grant T & Grigorieff N (2015) Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* **4**, 1–19.
141 Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV & Stark H (2015) Structure of the *E. coli* ribosome-EF–Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* **520**, 567–570.